SciDAC Visualization and Analytics Center for Enabling Technologies
Semi-Annual Progress Report
April 2008 through September 2008

E. Wes Bethel[*] and Chris Johnson[†]
Principal Investigators

Charles Hansen, Valerio Pascucci, Claudio Silva, Allen Sanderson [‡]

Sean Ahern, George Ostrouchov, Dave Pugmire, Jeremy Meredith[§]

Hank Childs, Peer-Timo Bremer,
Daniel Laney, Ajith Mascarenhas, Kathleen Bonnell, Brad Whitlock[¶]

Ken Joy, Christoph Garth, Bernd Hamann[||]

Gunther Weber, Prabhat, Oliver Rübel, Janet Jacobsen, Cecilia Aragon[**]

October 1, 2008

[*]Lawrence Berkeley National Laboratory
[†]Scientific Computing Institute, University of Utah
[‡]Scientific Computing Institute, University of Utah
[§]Oak Ridge National Laboratory
[¶]Lawrence Livermore National Laboratory
[||]University of California – Davis
[**]Lawrence Berkeley National Laboratory

Contents

1 Executive Summary

The SciDAC Visualization and Analytics Center for Enabling Technologies (VACET) focuses on leveraging scientific visualization and analytics software technology as an enabling technology for increasing scientific productivity and insight. Our mission is to foster scientific insight through creating and deploying effective data understanding technology that is truly responsive to the needs of our stakeholders in the scientific research community who are "awash in data." It is widely accepted that one of the bottlenecks in contemporary science is the need to gain insight from vast collections of complex data.

The vision for our Center is to respond directly to this challenge by adapting, extending, creating when necessary and deploying visualization and data understanding technologies for our science stakeholders. Organized as a Center for Enabling Technologies, we are well positioned to be responsive to the needs of a diverse set of scientific stakeholders in a coordinated fashion using a range of visualization, mathematics, statistics, computer and computational science and data management technologies.

We are pleased to report accomplishments during the period of April 2008 through September 2008, both in terms of impact for scientific stakeholders and in terms of providing leadership in the visualization and analysis community.

VACET has made a substantial impact on the SciDAC community:

- Accelerator Modeling. Our research has resulted in two major capabilities needed by accelerator scientists. First, we combined forces with researchers from the SciDAC SDM Center to implement high-performance query-driven visual data analysis in a production-quality, parallel capable software application (see Section 2.1). This capability replaces a software process that consumed hours of runtime with one that runs in seconds on the Cray XT4 at NERSC. Second, we developed a technique for automatic detection of particles undergoing wakefield acceleration (see Section 2.2). This new capability also replaces an older, manual process. Our stakeholder, C. Geddes (LBNL), reported these findings in his presentation at the SciDAC 2008 meeting in Seattle, WA.

- Adaptive Mesh Refinement Visualization. One significant accomplishment is that the SciDAC Applied Partial Differential Equations Center has adopted VisIt as its project-wide visual data analysis software application. Additionally, CCSE researchers, who are providing the simulation code to the SciDAC Community Astrophysics Consortium, have declared that VisIt is the visualization application its researchers should use. Therefore, VACET has achieved one of its primary mission objects: provide production-quality, parallel capable Adaptive Mesh Refinement visualization software to the DOE scientific community. This accomplishment helps to realize the vision for SciDAC: software infrastructure to effectively make use of parallel computing platforms for scientific knowledge discovery. Other accomplishments include: (1) performance optimizations so that time required for visualization operations on a reference APDEC dataset have been reduced by an order of magnitude; (2) support for "mapped AMR grids" are part of VisIt's production release. Our AMR visualization effort is described in Section 2.3.

- Climate. Our team is working with the SciDAC Science Application "Design and Testing of a Global Cloud-Resolving Model" to: (1) achieve high levels of I/O performance on the Cray XT4 system at NERSC; (2) design and implement an effective, multi-resolution data model; (3) develop and deploy the production-quality visual data analysis software infrastructure that will be used by their team. Early results have focused on testing and improving I/O

performance on `franklin.nersc.gov`, generating early visualizations to confirm the code and data I/O layers are functioning correctly, and preliminary analysis and design of a high-performance, multi-resolution data model. See Section 2.6.

- Climate. Our team has delivered production-quality 3D visualization capabilities that are now included with the ESG's Climate Data Analysis Toolkit (CDAT) (Section 2.7). Also, our team is exploring a technology that would help streamline and improve climate data analysis workflow management with ESG researchers (Section 2.8).

- Combustion. We have performed research and development of new topological analysis techniques that have proven very useful in performing quantitative analysis of large datasets produced by combustion simulations. Recent work describing topological analysis of a DNS combustion code (Section 2.5) was presented by the scientific stakeholder at the SciDAC 2008 Program Meeting. In that presentation, she indicated this new capability allowed her to see, for the first time ever, specific quantitative information about the evolution of combustion processes. This insight was not possible using traditional visualization tools (e.g., volume rendering). Related, toplogical analysis of AMR-based combustion models (Section 2.4) offers the first-ever insight into the relationship between the level of turbulence and its impact on the size and shape of combustion regions in a lean, premixed hydrogen flame simulation.

- Fusion. Our team members have been awarded a SAP project that formalizes the relationship with several different fusion projects. That work will focus on particle path and field visualization and analysis (Section 2.9), and will leverage prior VACET work and technology applied to both Fusion and Accelerator modeling projects. Additionally, our team members have fostered a close working relationship with Tech-X Corporation with the aim of delivering a production-quality, parallel-capable visualization application for use in all Tech-X fusion and accelerator projects (Section 2.10).

VACET continues to set the standard for visualization research productivity, outreach, and service:

- Publications: we report 28 peer-reviewed journal articles, 13 peer-reviewed works that appear in our field's leading conferences, several invited articles, book chapters, technical posters and technical reports (see Section 5.1.

- Invited presentations: we have delivered approximately twenty different invited presentations (Section 5.2).

- Software Tutorials: our team has delivered five tutorials on use and application of VACET software to scientific stakeholders (Section 5.3.)

- Workshop participation: VACET has contributed to five different domestic and international workshops on topics ranging from high energy fusion to mathematics for petascale data (Section 5.4).

- Awards. VACET wins three (of ten) "People's Choice Awards" at the SciDAC 2008 Program meeting in Seattle, WA (Section 5.5).

- Service. VACET researchers have served as technical reviewer for six different journals, conferences and funding programs; and served on the Program Committee (or as general chair or co-chair) for 26 different technical conferences and symposia (Section 5.6).

VACET also has been laying the groundwork for future results, both in terms of visualization research and in terms of deploying technology to our stakeholders:

- Streamlines. We have developed and deployed a parallel capable "streamlines engine" that will address needs from multiple science stakeholders (combustion, fusion, combustion, turbulence, astrophysics). This important new capability, described in Section 3.1, will help scientists gain deeper understanding into complex, time-varying and multi-grid vector field data produced by large-scale simulations. This work is an excellent example of VACET research, development and deployment being driven by science stakeholders' needs.

- Embedded Boundary/Material interfaces. Recent research has produced a highly accurate technique for computing embedded boundaries/material interfaces from simulation data containing cells with volume fraction data. The need to compute and display such interfaces is high on the priority list of one of our stakeholders, APDEC. While we have recently deployed legacy code that computes and displays embedded boundaries, we are working towards deploying the new Active Interface technique in VisIt, our production visualization application (see Section 4.3), and extending it to work with data on AMR grids.

- High Performance Query-Driven Visualization. As part of our research portfolio and in collaboration with the SciDAC SDM Center, we have developed a set of data structures and algorithms that show effective use of QDV on time-varying AMR datasets (Section 4.1). In collaboration with the SDM Center and the SciDAC Institute for Ultrascale Visualization, we have developed a novel approach for indexing suitable for use on highly multithreaded, GPU platforms (Section 4.2).

2 Specific Stakeholder Projects

2.1 Accelerator Modeling: High Performance Visual Data Analysis

Background and Motivation

Laser wakefield simulations model the behavior of individual particles as well as the behavior of the plasma electric and magnetic fields. Output from these simulations can become quite large: todays datasets, such as the ones we study here, can grow to be on the order of 200GB per timestep, with the simulation producing approximately 100 timesteps. The scientific challenge we help address in this study is first to quickly find particles that have undergone wakefield acceleration, then trace them through time to understand acceleration dynamics, and perform both visual and quantitative analysis on the set of accelerated particles.

The primary objective of this project is to explore a novel technique for rapid data exploration and subsetting operations to support the types of investigatory work patterns in use by accelerator scientists. One scientific impact of our work is that we have vastly reduced the duty cycle in visual data exploration and mining. In the past, accelerator scientists would perform the "trace backwards" step using scripts that performed a search at each timestep for a set of particles. Runtimes for this operation were on the order of hours. Using our implementation, those runtimes are reduced from hours to seconds.

Science Stakeholders and Collaborators

The science stakeholders on this project, listed below, are part of the SciDAC Community Petascale Project for Accelerator Science and Simulation (ComPASS) project, and are INCITE awardees for

cycles/storage at NERSC:

- Cameron G. R. Geddes and Estelle Cormier-Michel, LBNL.
- Peter Messmer, Tech-X Corporation.

In addition, we are collaborating with Kesheng Wu of the SciDAC Scientific Data Management Center on this project. Wu and the SDM Center have been instrumental in extending FastBit[1].

Accomplishments

1. **Data Modeling and Formats.** In order to apply FastBit-accelerated index/query operations to the laser wakefield simulation data, we used our HDF5-FastQuery technology (see `https://codeforge.lbl.gov/projects/h5part`), which provides a veneer API atop HDF5 and FastBit to implement simplified access to I/O and index/query capabilities. We also had to update the VisIt plugin that loads such data.

2. **Effective Visual Information Display.** We developed a new technique for displaying parallel coordinates plots of large datasets using 2D histograms (having uniform- or adaptively-spaced bins.)

3. **Effective Query Interface.** We performed the software engineering in VisIt necessary to "link" the parallel coordinates plot, which provides sliders on each parallel coordinates axis to indicate subset selection, with the database plugin to extract the desired data subset from the larger data file. VisIt then passes the extracted subset from the file loader to downstream processing and visualization components.

4. **Visual Display of Query Results.** We developed a methodology for effectively presenting query results within the context of the larger dataset (see Figure 1). The idea is to provide two types of information: (1) to visually indicate the query results within the context of the original scientific datasets; and (2) to visually indicate statistics about the query result, which we accomplish with multiple, overlayed parallel coordinates plots.

5. **Accelerated Particle Tracking.** Once the conditions associated with particle acceleration are identified, we devised the means for quickly finding all such particles across the many timesteps of data produced by simulation. Results of primary/secondary acceleration, 3D visualization, and particle tracking are shown in Figure 2.

6. **Performance Experiment.** There are two main operations that comprise the HPC visual data analysis operation: (1) computing conditional 2D histograms for visual information display as well as the query interface; (2) particle tracking across the many timesteps produced by simulation. Our performance experiment uses a source dataset that is approximately 1.5TB in size, including the index data. We measure parallel performance and scalability of the histogram computation and particle tracking operations (see Figure 3).

7. **Publications and Outreach.**
   - Paper accepted to technical program at SC08:O. Rübel, Prabhat, K. Wu, H. Childs, J. Meredith, C.G.R. Geddes, E. Cormier-Michel, S. Ahern, G.H. Weber, P. Messmer, H. Hagen, B. Hamann and E.W. Bethel. High Performance Multivariate Visual Data Exploration for Extremely Large Data. SC08, Austin TX, November, 2008.

[1] FastBit is software that implements high performance, compressed bitmap indexing. It won an R&D 100 award this past year. More information is located at `http://sdm.lbl.gov/fastbit`.

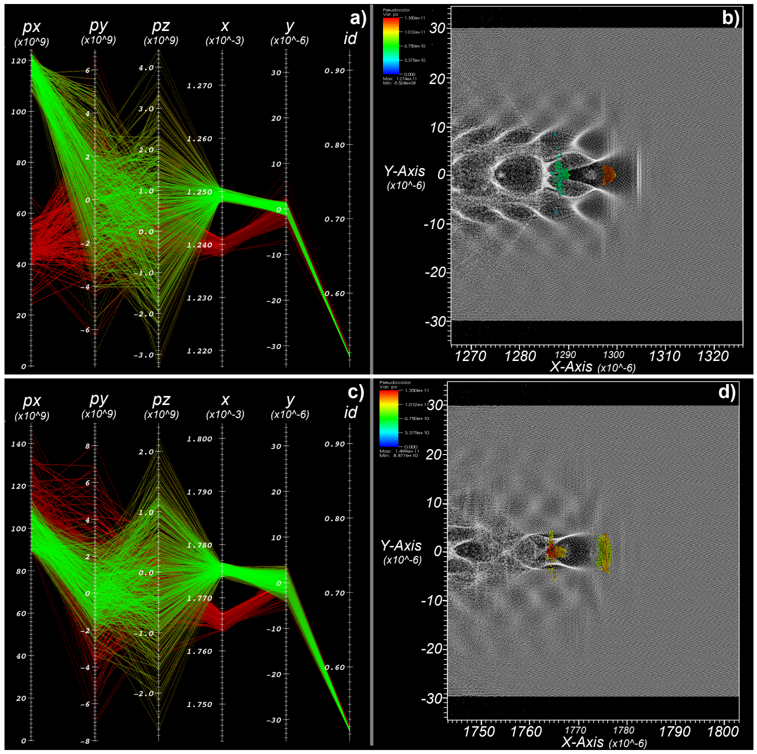


Figure 1: a) Parallel coordinates and b) pseudocolor plot of the beam at $t = 27$. Corresponding plots c,d) at $t = 37$. The context plot, shown in red, shows both beams selected by the user after applying a threshold of $px > 8.872 * 10^{10}$ at $t = 37$. The focus plot, shown in green, indicates the first beam that is following the laser pulse. In the pseudocolor plots b) and d), we show all particles in gray and the selected beams using spheres colored according to the particle's x-momentum, px. The focus beam is the rightmost bunch in these images. At timestep $t = 27$, the particles of the first beam (green in figure a) show much higher acceleration and a much lower energy spread (indicated via px) than the particles of the second beam. At later times, the lower momentum of the first beam indicates it has outrun the wave and moved into decelerating phase, e.g at timestep $t = 37$.

- Poster at IEEE Visualization 2008. O. Rübel, Prabhat, K. Wu, H. Childs, J. Meredith, C.G.R. Geddes, E. Cormier-Michel, S. Ahern, G.H. Weber, P. Messmer, H. Hagen, B. Hamann and E.W. Bethel. Application of High-performance Visual Analysis Methods to Laser Wakefield Particle Acceleration Data." Poster at IEEE Visualization 2008, Columbus, Ohio, October 19-24, 2008.
- SciDAC 2008 presentation. Geddes presented recent research results of his laser-wakefield project. The results included visual data analysis (images, movies) made possible by our contributions, as well as statements about how this new capability enables scientific knowledge discovery.

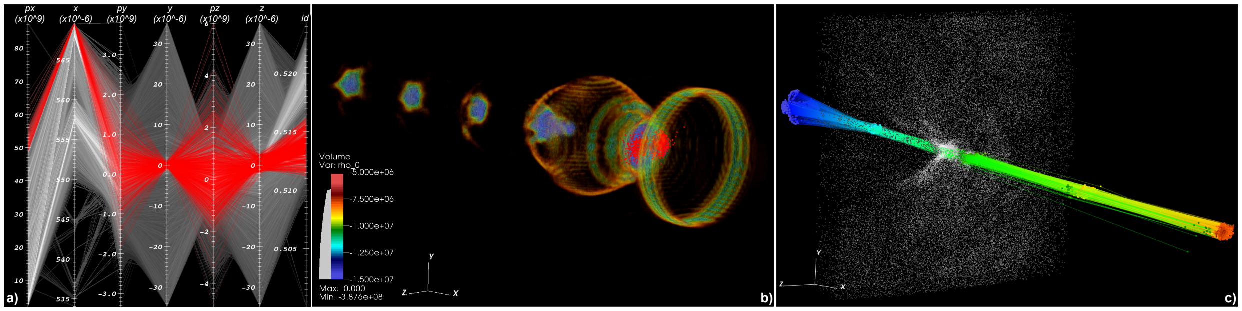
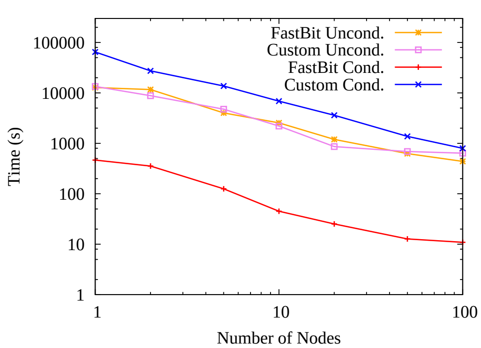
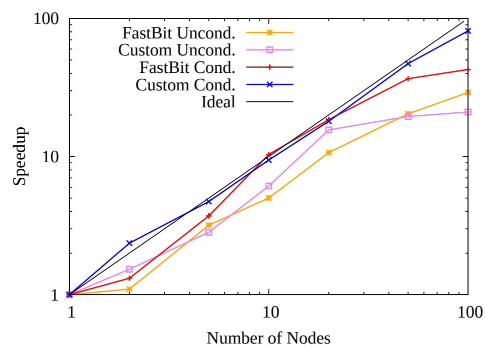


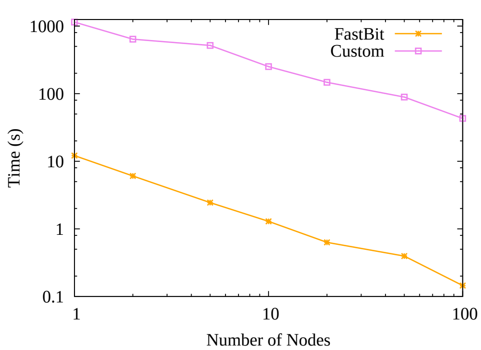
Figure 2: a) Parallel coordinates of timestep $t = 12$ of the 3D dataset. Context view (gray) shows particles selected with $px > 2 * 10^9$. The focus view (red) shows particles satisfying the condition $(px > 4.856 * 10^{10})$ && $(x > 5.649 * 10^{-4})$, which form a compact beam in the first wake period following the laser pulse. b) Volume rendering of the plasma density and the selected focus particles (red). c) Traces of the beam. We selected particles at timestep $t = 12$, then traced the particles back in time to timestep $t = 9$ when most of the selected particles entered the simulation window. We also traced the particles forward in time to timestep $t = 14$. Color indicates px. In addition to the traces and the position of the particles, we also show the context particles at timestep $t = 12$ in gray to illustrate where the original selection was performed. We can see that the selected particles are constantly accelerated over time (increase in px).



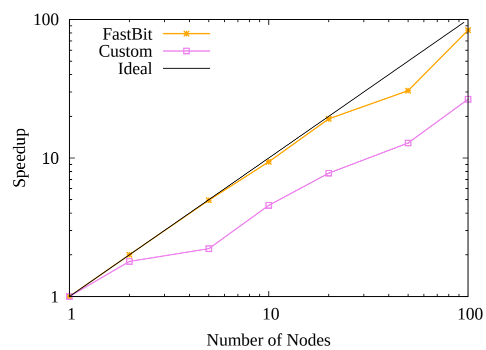
(a) Timings for parallel computation of histograms.



(b) Scalability of parallel histogram computation.



(c) Timings for parallel particle tracking.



(d) Scalability of parallel particle tracking.

Figure 3: These results show performance and scalability of parallel conditional histogram computation and temporal particle tracking. The science impact of this performance result is that the process of tracking particles undergoing acceleration across many timesteps of multi-terabyte datasets, a process that once took hours, now is accomplished in seconds. This capability is deployed in a production, quality visual data analysis application, and is readily applicable to other science domains.

2.2 Accelerator Modeling: Using Analysis to Identify Beam Particles

The scientific stakeholders on this effort[2] are both experimentalists and computational scientists. The primary objective for this project is to devise a mechanism for automatically finding particles that are undergoing acceleration in a laser-wakefield simulation dataset. This capability would help to automate a process that is currently performed manually. In the long run, this capability will certainly help to accelerate data understanding. As it evolves, this technique will likely be integrated with other tools (e.g., high performance visual data analysis) as part of a broad set of HPC tools for scientific knowledge discovery in accelerator science.

Accomplishments

We devised a methodology for automatically identifying beam particles, ie., those undergoing wakefield acceleration, from data produced simulation:

- The methodology identifies "bunches" of particles likely to be part of the beam as those having both high momentum and are in close spatial proximity.
- For each bunch of particles from all timesteps' worth of data, we use a graph algorithm to track bunch movement and evolution across timesteps.
- Separately, we use fuzzy clustering to classify particles as either "beam" or "non-beam," where classification is continuous rather than discrete.
- Results of fuzzy clustering are compared with those from the space/momentum classification stage. Where there is agreement between these two models, we have a high quality beam.

Future Work

- Further work in refining our ability to perform unsupervised classification of beam particles.
- Explore use of parallel analysis techniques, e.g., Parallel R in conjunction with our colleagues at ORNL.
- Since the work thus far focuses exclusively on particles, and ignores the other available data (e.g, electric field), longer term work will focus on devising the methodology to better understand the relationship between particles undergoing acceleration and their surrounding environment.

2.3 Adaptive Mesh Refinement Visualization

Both our primary stakeholder teams – APDEC[3] and CCSE[4] – have a need for production-quality AMR visual data analysis infrastructure. Both teams have a separate in-house tools they have been using and funding over the years. Both teams have expressed the urgent need to get out of the business of developing and maintaining such software infrastructure. Additionally, both teams have expressed the need for new capabilities that lie outside the scope of what is possible with their

[2] C. Geddes and E. Cormier-Michel (LBNL), P. Messmer (Tech-X) are INCITE awardees at NERSC and part of the SciDAC COMPASS project.

[3] The SciDAC Applied Partial Differential Equations Center, `http://www.apdec.org`

[4] The LBNL Center for Computational Sciences and Engineering, `http://seesar.lbl.gov/CCSE/Research/index.html`

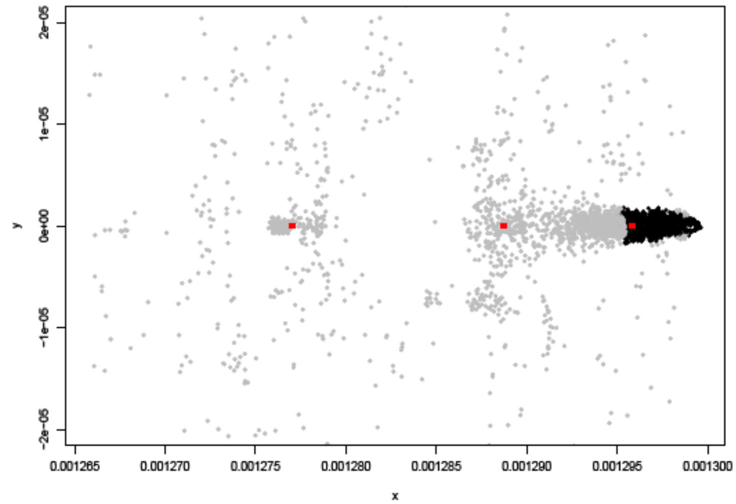


Figure 4: Beam-point candidates combined with fuzzy clustering. Beam points, which are comprised of high-momentum particles having spatial coherence are shown in red. The results of fuzzy clustering shows beam particle candidates in black, and non-beam particles in gray.

existing technology. The activities to support these teams can be described with five high level thrusts:

1. Invest in critical, missing visualization algorithms.
2. Planned interface changes to ease the transition of their user community.
3. Planned functionality improvements specific to APDEC.
4. Maintenance of software infrastructure, bug fixes, and other short term, unforeseen changes.
5. Performance improvements specific to AMR data sets.

A more detailed list of specific needs is documented on the VACET twiki[5]. That list contains around forty different features requested by customers from APDEC and the Astrophysics SAP project. Because the tasks are so numerous and since many will take a significant amount of time to accomplish, our approach is to prioritize these features to optimize the intersection between (1) the stakeholder's assessment of the features importance and (2) the degree of difficulty in implementing the feature. With this prioritization, we order our work to achieve impact quickly while continuing work on more difficult and challenging research and engineering problems. The tasks of each of the four categories are itemized below along with labor efforts, stakeholder priority, and status (in some cases where appropriate).

VACET Team

- LBNL: W. Bethel, G. Weber (Stakeholder Project Manager).
- LLNL: H. Childs, B. Whitlock, K. Bonnell.

[5] `http://www.sci.utah.edu/vacetwiki/index.php/Collab:APDEC_VisIt`

- ORNL: J. Meredith.

- UC Davis: K. Joy, J. Anderson.

- Related: VACET Software Engineering team, for critical infrastructure technology, like high-quality volume rendering and streamlines.

Stakeholders

- Phil Colella, LBNL and the SciDAC Applied Partial Differential Equations Center (APDEC).

- John Bell, LBNL and Ann Almgren, LBNL (CCSE).

- Stan Woosley (SciDAC Science Application: Computational Astrophysics Consortium: Supernovae, Gamma Ray Bursts, and Nucleosynthesis).

- Other beneficiaries
  - Mike Barad, Environmental Fluid Mechanics Laboratory, Stanford University.
  - Ravi Samtaney, Theory Department, PPPL (Fusion).

Accomplishments this Period

While a number of individual accomplishments are described below, perhaps the single most significant accomplishment is that APDEC has adopted VisIt as its project-wide visual data analysis software application. Additionally, CCSE researchers, who are providing the simulation code to the SciDAC Community Astrophysics Consortium, have declared that VisIt is the visualization application its researchers should use. Therefore, VACET has achieved one of its primary mission objects: provide production-quality, parallel capable Adaptive Mesh Refinement visualization software to the DOE scientific community. This accomplishment helps to realize the vision for SciDAC: software infrastructure to effectively make use of parallel computing platforms for scientific knowledge discovery.

During this period, we have performed a substantial amount of software engineering aimed at meeting a number of stakeholder needs as well as bug-fixes. These accomplishments, while not the type of result that would merit publication in a journal or conference, are absolutely critical to this project's success.

The labels (B7), etc., indicate the associated need/task on internal project management task lists.

- (E4) Tune for large numbers of patches. This accomplishment is one of the most significant developments of the last six months. Described further in its own section below.

- (B7) Repicking with new variables.

- (B10) Creating spreadsheets for existing picks.

- (B11) Convenience macros for Chombo to VisIt transition.

- (C4) Mapped grids.

- (C10) Calculate the Laplacian using a single layer of ghost zones.

- (D6) Networking bug (believed to be resolved).

- (D10) Support for p4_ch driver.
- (D11) Clean up runaway engines when the client disconnects abnormally.
- (E1) Performance of the Subset plot.

Description of Tuning Effort

First, note that the tuning described here was for a specific use case and more tuning has been requested for other use cases. Still, the work done here is general and should benefit the other use cases.

Even when production quality visualization tools contain the capabilities that stakeholders need, these tools can fail because they are not tuned for the real world data sets that customers generate. In the work described here, scientists from the Applied Partial Differential Equations Center (APDEC) of the Scientific Discovery through Advanced Computing (SciDAC) program of the Department of Energy (DOE) were unable to use the open source visualization and analysis tool VisIt to study some of the simulations they were performing. Although VisIt performs well on most large data sets, the data created by APDEC scientists was somewhat unique to VisIt and required tuning. So scientists from the Visualization and Analytics Center for Enabling Technologies (VACET) studied and improved various bottlenecks in algorithms used by VisIt. For the data set in question, this resulted in a factor of ten improvement, from 811 seconds to 72 seconds, an improvement that took the tool from being unusable to usable. Further, the performance quoted is for serial processing and does not take into account optimizations such as I/O caching, which would make visualization and analysis activities even more interactive. Also, it is worth noting the optimizations described here have been contributed back to VisIt and will benefit the entire community of VisIt users.

The data set produced by the APDEC center was an Adaptive Mesh Refinement (AMR) data set containing 55,000 patches spread across six refinement levels. The size of the file was relatively modest (1.5 GB), as was the number of cells (approximately 15 million). The reason this data set stressed VisIt performance was because VisIt handles each patch as its own data set, meaning that it had to handle 55,000 "pieces" of data. By way of contrast, many data sets that have billions of cells will be decomposed into as few as 1,000 pieces, meaning that VisIt will have many fewer pieces to process. Since VisIt had not been stressed in this dimension before, several $O(n^2)$ algorithms were uncovered and fixed.

The optimizations to improve performance on this data set are as follows:

- VisIt removes the coarse portions of the AMR grid that are refined by finer levels. It does this with the help of a data structure that described the "nesting" between levels. The file format reader was using an algorithm to determine overlap between patches by comparing the spatial extents of each patch with every patch in the immediately coarser refinement level. We sped this up by implementing an interval-tree based approach that dropped the time from 11.25 seconds to 0.25 seconds.
- VisIt contains a "transform manager" that checks to make sure that all data is suitable for visualization. This module contained a routine for identifying patches that walked the list of patches unnecessarily, resulting in an $O(n^2)$ algorithm, where n is the number of patches. For the majority of VisIt visualizations (which have much less than 1000 pieces of data) this slow algorithm rarely affected overall performance. For the APDEC data, however, the transform manager was taking 360 seconds. We improved the mechanism for identifying patches and the run time for this routine dropped to 2.2 seconds.

Operation	Before (seconds)	After (seconds)
Nesting calculation	11.25	0.25
Transform manager	360.00	2.20
Applying nesting	0.56	0.14
Removing pieces	14.00	0.35
Memory management	420.00	46.00
Total	811.00	72.00

Table 1: Summary of performance improvements.

- When removing the portions of a patch that are refined out, VisIt was repeatedly calculating the same meta-data. Calculating this information once and caching it resulted in a speedup of 4X, from 0.56 seconds to 0.14 seconds.
- We identified another algorithm for removing data pieces that wont affect the result which was iterating in an $O(n^2)$ way over the patches. Fixing this caused a speedup from 14 seconds to 0.35 seconds.
- The largest improvement came from memory management. For each algorithm, which is applied to each patch, there are many allocations and deallocations. Worse, some small pieces of memory are allocated and not deallocated for long periods of time, effectively fragmenting the memory. The memory manager on Linux handles extreme fragmenting poorly and performance dropped significantly as execution proceeded. (We were able to conclude that memory fragmentation was the problem by creating a standalone program that only did allocations and deallocations and which mimicked the patterns of VisIt. This program reproduced the problem and was ultimately sent to RedHat Linux for further study.) We were able to overcome the memory management slowness by incorporating an open source library from "Google perftools" named `tcmalloc`, a replacement memory management scheme that performs better with memory fragmentation. Simply by linking in `tcmalloc`, all processing performance (that is, non-I/O performance) improved from 420 seconds to 46 seconds. We modified VisIts build system to now include `tcmalloc` and future distributions of VisIt will include this library.

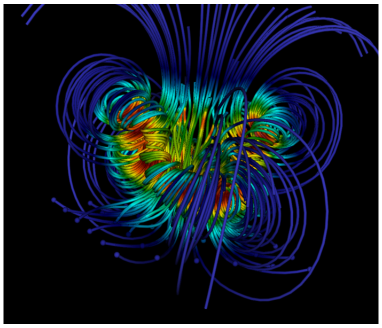
Future Work

High Priority Items:

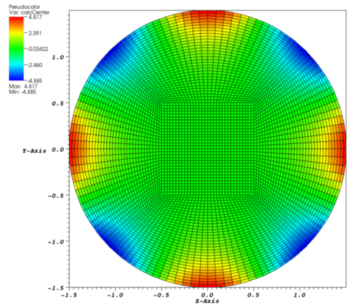
- Embedded boundary support. A separate but coordinated effort within VACET is focusing on this problem (see Section 4.3.
- Streamlines for AMR data. At present, streamlines do not descend into finer patches/boxes of the hierarchy. Full AMR streamline support is work in progress (see Section 3.1). We expect to have an initial implementation by the next progress report in the Spring of 2009.
- Double precision support in pipelines.

Priority Items:

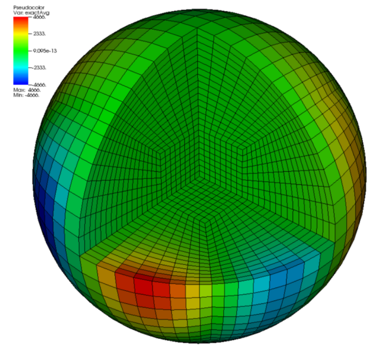
- Performance improvements for EB data (E3).



(a) Streamlines visualization of two vortex cores merging. Image produced by Dave Pugmire (ORNL) using AMR data produced by APDEC's Chombo code.



(b) Psuedocolor plot of a 2D mapped AMR grid.



(c) Psuedocolor plot of a 3D mapped AMR grid.

Figure 5: Accomplishments in this period include: support for a new streamlines algorithm (left image), as well as 2D (middle) and 3D (right) mapped AMR grids.

- Performance improvements for mesh plots (E2).
- Scalable selection of patches (C9) .

Need priority re-evaluation (due to limited VACET resources):

- Animating Slices (B5).
- Automatic repicks (B7).
- View Interactors (B8).
- (C4) with new file format, uncertain whether there will be a new file format. However, once we have the two-file solution going to a single EB file format should be relatively low effort.

Risks

- The APDEC team is transitioning to VisIt and that will lead to: (1) increased support costs and (2) many additional requests for features and bugs. These additional requests will likely take precedence over some of the planned tasks.
- The tasks for APDEC only total approximately 12 weeks. The personnel assigned to work on APDEC work are also pursuing A6, A8, A10, and A11, each of which is a high priority item with a large cost. There is a high likelihood of insufficient resources within VACET to meet these objectives.

2.4 Combustion: Structural Analysis of AMR Combustion Simulation Data

Our team has been working with John Bell (LBNL) and researchers at LBNL's Center for Computational Sciences and Engineering (CCSE) to perform research, development, and application of topological analysis techniques to provide new insights into the combustion process. Bell's group works with large AMR 3D time-varying combustion simulation datasets, and needs analysis tools

to aid their understanding. They are interested in formation and evolution of “extinction zones.” They may be able to provide us with a boolean volume of ”burning” and ”non-burning” regions, and we will perform a detailed analysis of the structural characteristics of the network of low temperature regions. They are also interested in using topological methods to identify those zones with topology-based methods. To this end we will use topological tools both feature characterization and for robust time tracking.

The Team

VACET

- Valerio Pascucci (Utah), Team Leader.
- Peer-Timo Bremer (LLNL), topological feature extraction.
- Gunther Weber (LBNL) data management and topology-based analysis.

CCSE – Science Stakeholders

- John Bell and Marc Day (LBNL), Center for Computational Sciences and Engineering.

Accomplishments

We have expanded and finalized the topology based analysis of three simulations of lean hydrogen flames under different levels of turbulence. Understanding combustion processes over a broad range of operational regimes is of great interest for a variety of applications such as engine or power plant design. To this end, there has been considerable recent interest in the development of premixed burners capable of stably burning ultra-lean hydrogen-air fuel mixtures. Such burners could, for example, be used as one component of a clean-coal power plant utilizing hydrogen extracted from coal gasification. Lean premixed systems are subject to a variety of hydrodynamic and combustion instabilities that render practical flame stabilization, and traditional approaches to flame analysis, extremely difficult. The flames burn in a cellular mode that is highly nonuniform, time-dependent, and difficult to characterize. The stake holders are interested in understanding how many independent burning cells exist at any one time, their areas, and evolution over time. In particular the focus is on comparing the flames at different turbulence levels.

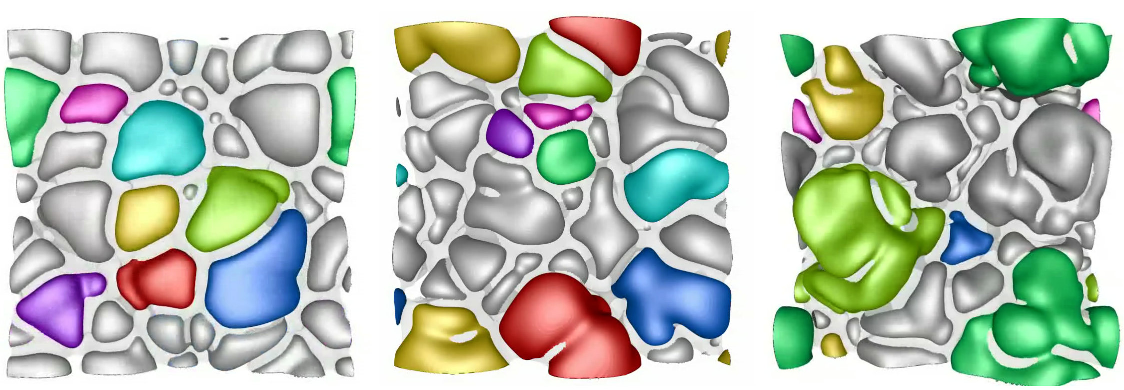


Figure 6: These images show the results of topological analysis that segment data into regions of combustion. The source data are no turbulence (left image), weak turbulence (middle), and strong turbulence (right).

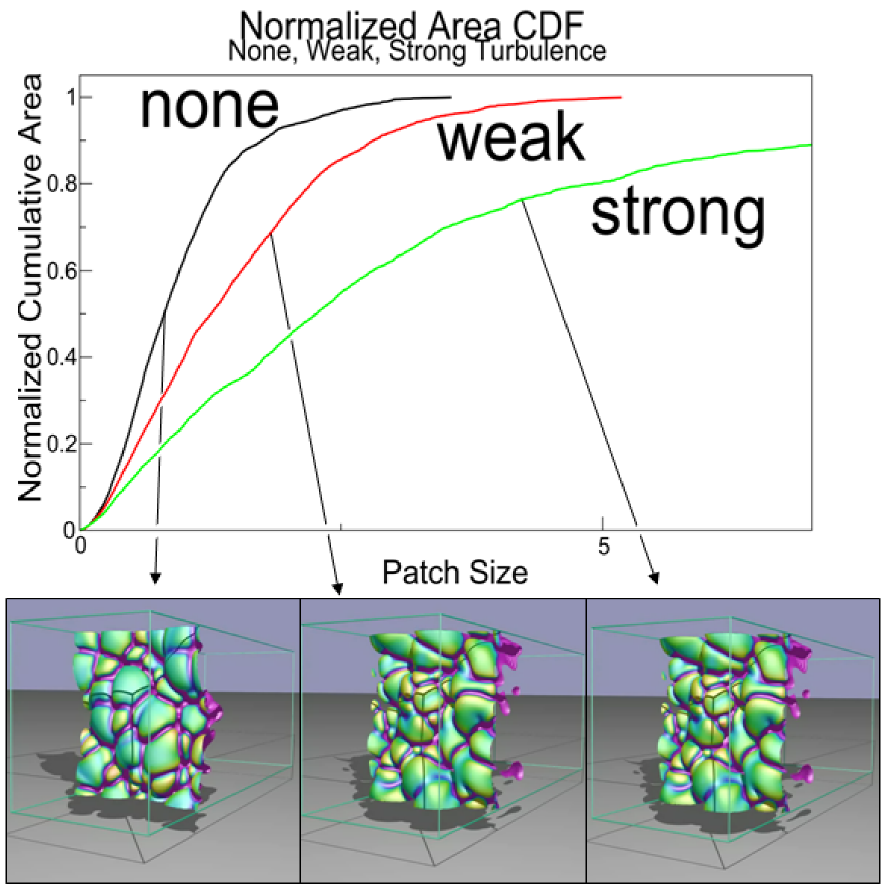


Figure 7: The topological analysis provides the means for quantitative comparison of combustion region attributes in the presence of no, weak, and strong turbulence.

We track all burning cells in all three simulations and create tracking graphs that encode the temporal evolution of all cells (Figure 8. The graph provides the stakeholders with in-depth information about how individual cells behave over time. The graph is used in multiple ways: First, by propagating colors along edges of the graph we can render color coded animations to show the evolution of certain cells. Second, the graph can be used to study interaction between cells on a individual scale and is also used for debug the segmentation algorithms. Finally, we use the graph to compare different spatial and temporal resolutions of each simulation to tease out similarities and differences.

Figure 8 shows the results of one such analysis, where we see a small portion of the tracking graph for the turbulence free case. Round nodes indicate burning cells segmented from an actual flame surfaces using the Morse complex. The numbers inside the nodes and along the branches indicate the identifier assigned to this particular cell in the segmentation. The diamond shapes represent topological events in between time steps and thus have no identifier attached. To provide better navigation we use the space inside the diamonds to indicate the simulation time. Red/green nodes and diamonds signal a merge/split event and turquoise structures a birth or death event. Using the renderings above and below the graph one can follow the event chain.

Between time 550 and 552.5 the purple cell 4 merges with cell 18 via a small connection across

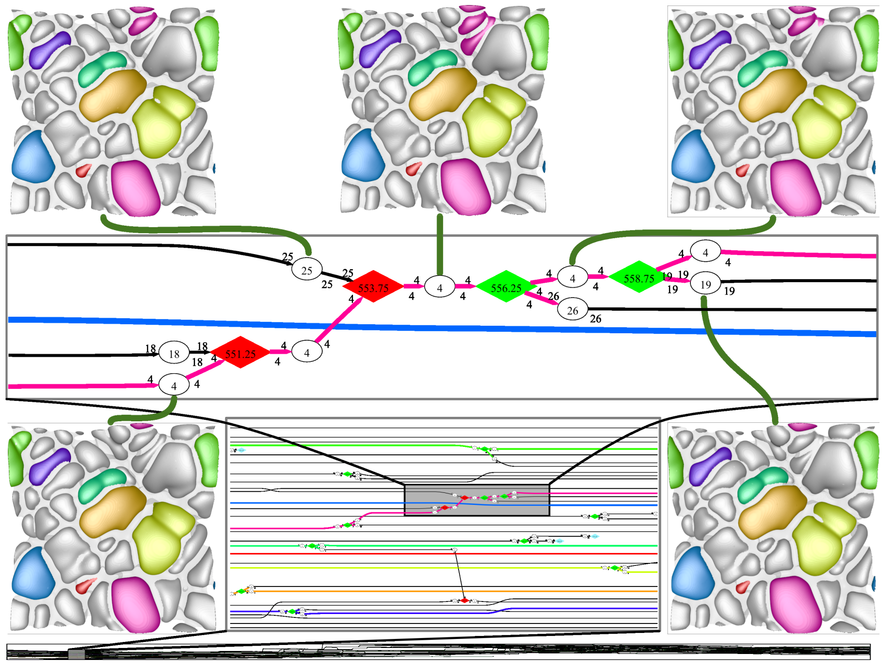


Figure 8: Detailed graph analysis applied to the turbulence-free dataset. It provides detailed information about the evolution of burning cells/regions, including split/merge operations that would not be possible with traditional analysis methods.

the lower (periodic) boundary. Subsequently, what used to be cell 18 develops a connection with cell 25 resulting at time 555 in the three purple areas all connected via small bridges to a single cell. At time 557.5 the cell formerly labeled 25 has split leaving a small portion attached to cell 4 and creating a new cell 26. Finally, at time 560 the remains of cells formerly labeled 18 and 25 have split off cell 4 forming a separate cell 19.

In this period, we have submitted two separate publications describing different aspects of this work to two different venues/journals.

Future Work

Near Term

- Develop new streaming segmentation algorithm capable of dealing with three-dimensional features rather than features on surfaces.
- Implement the new segmentation algorithm.
- Expand two-dimensional analysis to analyze the genus of non-burning regions.

Longer Term

- Adapt the tracking algorithm to three-dimensional features.
- Apply the new pipeline to a larger, more complicated data of interest to the stakeholders.
- Compare results from two-dimensional and three-dimensional techniques.

2.5 Combustion: Quantitative Analysis of DNS Combustion Simulations

Our stakeholder, Jacqueline Chen at SNL-CA, has the need of being able to perform quantitative analysis of data produced by DNS combustion simulations to aid in the understanding of the combustion process. Our approach is to perform topological characterization of combustion features in DNS-produced data. The work has a particular focus on developing new insight about the genesis and evolution of "extinction pockets." Understanding how extinction and reignition happens in premixed hydrogen combustion has the potential of allowing better design of engines and power plants. We focus on the use of robust topological methods for segmentation and tracking of regions of high scalar dissipation rate that provide a good first order approximation of true extinction regions.

The Team

VACET

- Valerio Pascucci (Utah), Team Leader.
- Peer-Timo Bremer (LLNL), topological feature extraction.
- Dan Laney (LLNL), data management and streaming.
- Ajith Mascarenhas (LLNL), feature tracking.

Science Stakeholders

- Jacqueline Chen (SNL-CA).

Accomplishments

Multi-scale Segmentation of Dissipation Rate.

We completed the computation of multi-scale segmentation of regions high scalar dissipation rate. Worked with stakeholders at Sandia to develop a new S3D library and interface layer to run segmentation code. It features: (1) serial streaming mode with minimal memory and hardware requirements; (2) parallel mode coupled to S3D with all derived variables and features of the terascale DNS code available.

To develop the first complete, time analysis of high scalar dissipation rate regions, we visualize these regions in 3D and allow the scientists to change the region selection parameter in real time. Figure 9 shows the stoichiometric mixture fraction surface superimposed with the selected high scalar dissipation rate regions.

Correspondence Between Features and Combustion

- Compare quantities averaged over the portion of the stoichiometric mixture fraction surface inside and outside the features, normalized by the average over the entire surface.
- Comparison for all features intersecting the surface shows only small differences; average over portion of surface outside the features closely approximates the average over the entire surface.
- Significant change in conditional reaction with conditional scalar dissipation within feature suggests that not all features have high enough scalar dissipation to experience extinction.
- Comparison including only those features with large conditional chi shows much larger effect.

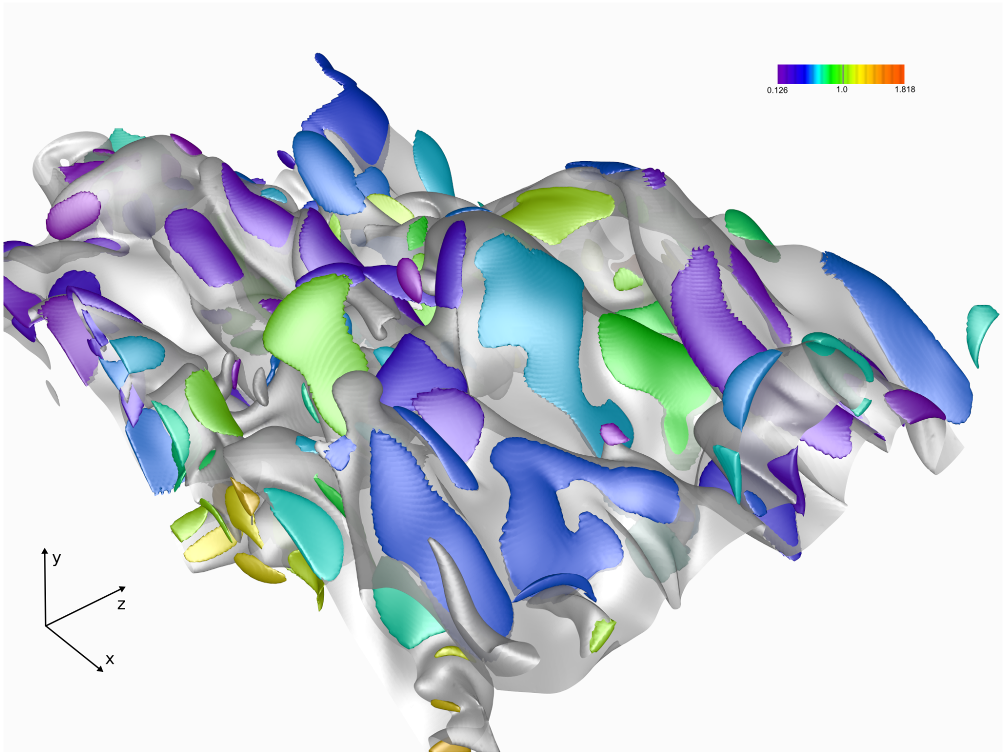


Figure 9: This image shows the stoichiometric mixture fraction surface superimposed with the selected high scalar dissipation rate regions.

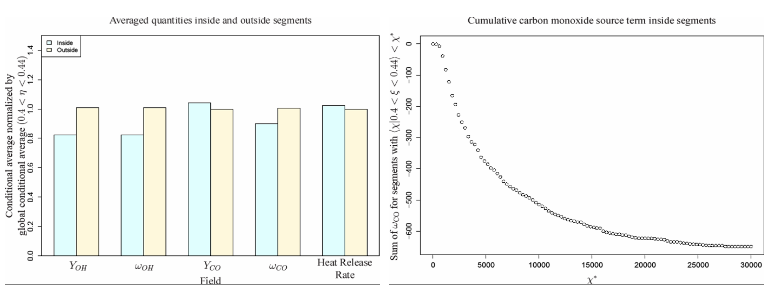


Figure 10: Comparison for all features intersecting the surface shows only small differences; average over portion of surface outside the features closely approximates the average over the entire surface.

Temporal History

- Connecting features with spatial/temporal overlap allows time tracking.
- Following the ten largest features at 20 jet times (vertical line) forward and backward in time produces the "tracking graph" (Figure 12).
- These features are the result of merges earlier in time.
- Later in time, the selected features split.
- Prior to 20 jet times, some of these features have been part of interacting families.

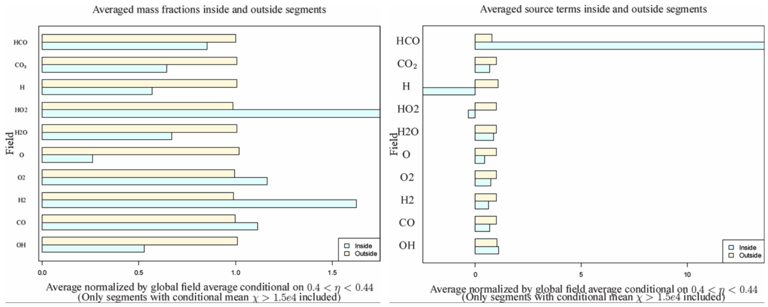


Figure 11: Comparison including only those features with large conditional chi shows much larger effect.

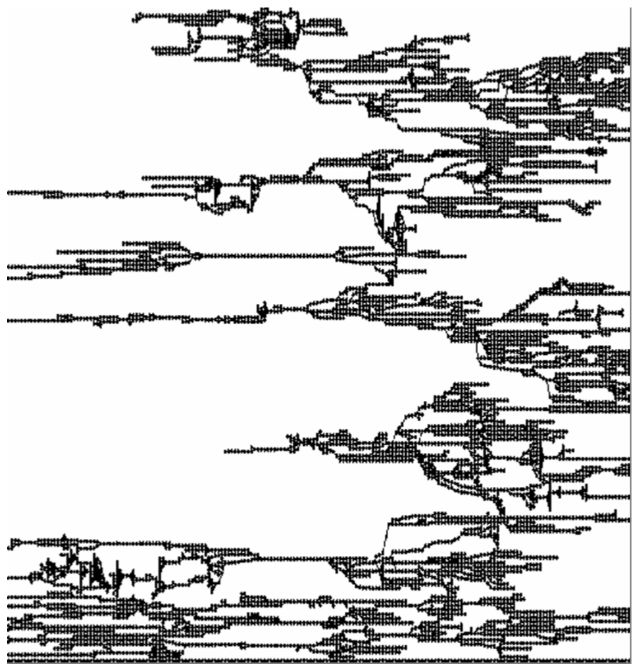


Figure 12: Connecting features with spatial/temporal overlap allows time tracking. Here, connected features are shown in graph form.

- The scalar dissipation rate magnitude for the tracking families fluctuates significantly.
- The fluctuations have a longer timescale for the families from this sample which have experienced the largest scalar dissipation rates (Figure 13).

Alignment between strain field and scalar gradient For non-reacting flows. The scalar gradient is preferentially aligned with the most compressive turbulent strain. Within the features, the preferential alignment is more pronounced than outside (Figure 14).

Impact.

Scientists were able for the first time to use a full 3D segmentation of extinction regions and follow in time their individual evolution, with detailed creation, destruction, merge and split events. This will allow a better understanding of complex dynamics of turbulent combustion processes and ultimately allow the design of cleaner and more efficient engines.

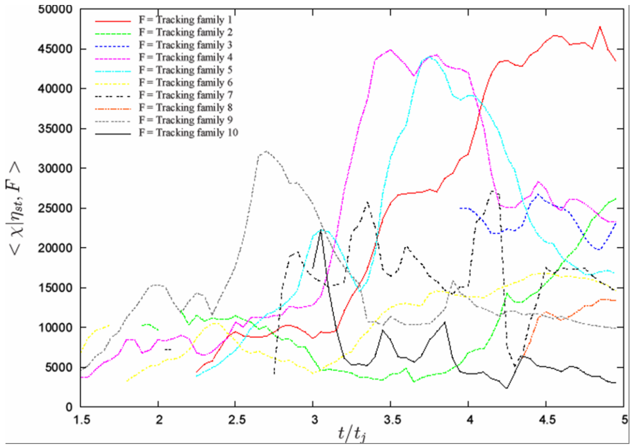


Figure 13: Scalar dissipation for tracked features.

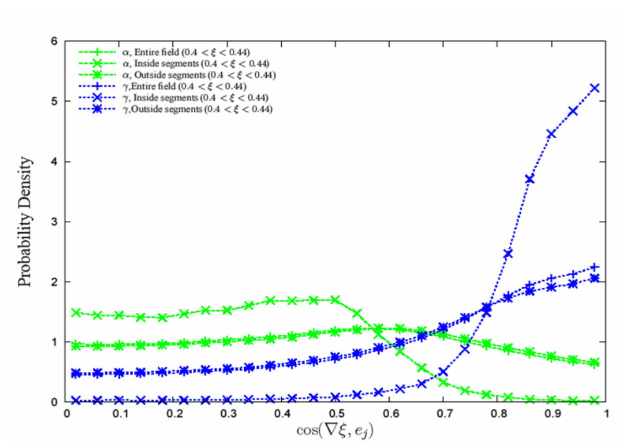


Figure 14: For non-reacting flows, the scalar gradient is preferentially aligned with the most compressive turbulent strain. Within the features, the preferential alignment is more pronounced than outside.

At the SciDAC 2008 program meeting, Jackie Chen's invited presentation included discussion that she was now seeing new features in her data never before possible. It is clear that this work is providing the ability for new scientific knowledge discovery.

Future Work

Continue collaboration with stakeholder to prepare paper analyzing the combustion aspects of the research.

2.6 Climate: Design and Testing of a Global Cloud-Resolving Model

A collaborative effort, which includes persons from VACET and the NERSC Analytics program, is working with members of the SciDAC Science Application "Design and Testing of a Global Cloud-Resolving Model." Their objective is to run a large-scale global cloud resolving model simulation at a high level of parallelism on the Cray XT4 system at NERSC (under an INCITE award). The simulation is expected to run on 20K+ cores and dump 1TB/hr. Our work in supporting

the scientists is threefold: (1) We are debugging and optimizing the collective IO performance on `franklin.nersc.gov`; (2) we are providing expert advice on the data model currently being proposed for writing the simulation output; (3) we are supporting their visualization/analysis needs by writing a custom plugin in VisIt to load their data directly without the need for format conversion.

Team Members

- VACET: Prabhat (Team Leader), Janet Jacobsen, Wes Bethel (LBNL).
- Stakeholders:
  - Dave Randall and Ross Heikes (CSU).
  - Karen Schuchardt, Bruce Palmer, and Annette Koontz (PNNL).

Stakeholder Needs

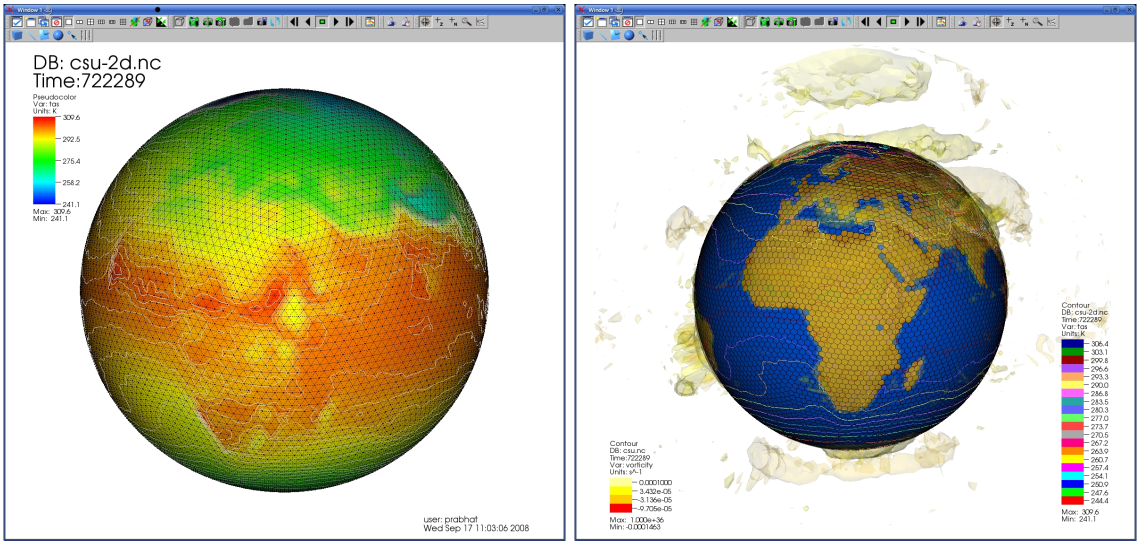
Efficient Collective I/O on franklin. Our stakeholders plan to run the GCRM simulations on 20K+ cores of franklin. Ideally, they want to do collective I/O, i.e. write to a single shared NetCDF file from multiple cores. Collective I/O will greatly simplify post-processing and analysis. They would like the I/O overhead to be 5% of the total time allocation, which amounts to a sustained IO performance requirement of approximately 2GB/s. Currently they are observing a collective IO rate of 100-700MB/s. There are a number of issues plaguing collective I/O performance on franklin, and we are working through multiple levels of the the system hardware/software stack to identify bottlenecks. The desired goal is to identify a well-tuned set of system parameters/configurations and thereby accelerate the GCRM IO code.

Data Model. Our PNNL collaborators are designing a richly featured, high-performance data model for the simulation output. There is a pressing need to integrate the I/O API with simulation code and get something working before the current year's NERSC allocation (INCITE award) runs out. We are providing them with consulting services on identifying a range of mid/long term functionality and performance issues. For instance, we are bringing their attention to current NetCDF-3 limitations, NetCDF-4/HDF5/H5Part as other alternatives, high-performance layout of data on disk (such as Morton-ordered space filling curves), metadata/layout conventions for NetCDF variables, etc.

Visualization and Analysis support for GCRM data. GCRM simulations are conducted on an icosahedral grid. The resulting NetCDF datasets (1 file per timestep) will likely be 10-100GBs in size and the aggregate dataset size is expected to be 100s of TBs. Conventional tools currently in use by scientists are not expected to be able to handle interactive analysis/vis of these datasets. Our plan is to develop a custom VisIt plugin for the icosahedral mesh and thereby utilize VisIt's rich set of vis/analysis features as well as parallel rendering infrastructure.

Accomplishments this Period

- **I/O Troubleshooting.** We conducted a large number of I/O tests using the IOR benchmark to analyze the effect of the following parameters: (1) I/O patterns; (2) blocking/transfer sizes; (3) Lustre filesystem parameters; (4) MPI-IO hints; (5) the number of I/O nodes; (6) number of OSTs; (7) MPI-IO and POSIX/HDF5/NetCDF interactions.
- We were able to confirm that Lustre's collective I/O performance lags far behind file-per-process I/O performance. A number of limitations were identified with Cray's `mpich` imple-



(a) VisIt plot of a 2D icosahedral mesh and surface temperature.

(b) VisIt plot of a 3D icosahedral mesh, land cover, surface temperature, and atmospheric vorticity.

Figure 15: Visualization of the climate simulation data on an icosahedral mesh. The climate simulation code, which is part of the SciDAC Climate Science Application, was run on the Cray XT4 system at NERSC. Its aim is to produce a global atmospheric circulation model with a grid-cell spacing of approximately 3 km, capable of simulating the circulations associated with large convective clouds.

mentation. A prototype `romio` implementation (OPAL) from the SciDAC SDM center was examined, though it did not offer significant performance improvements.

- Currently, the best strategy seems to be a careful selection of file striping, number of I/O nodes and transfer size. Careful tuning will result in collective I/O operations under the expected workload having the same I/O pattern and performance as in the file-per-process case.

- **Visualization and Analysis.**
  - We developed a prototype serial VisIt plugin to load the icosahedral mesh data from NetCDF files.
  - Based on preliminary feedback from scientists, custom visualizations were developed to demonstrate VisIt's functionality (see Figure 15).
  - Online tutorials were developed to facilitate ease-of-use and adoption for our collaborators (see `http://vis.lbl.gov/~prabhat/Incite19/`).

Future Goals

- Parallelization of the VisIt data loader plugin.
- Confirm that 64-bit integers mesh/variable data works in VisIt.
- Support for edge-centered and face-centered data for 3D meshes.
- Interactions w/ PNNL collaborators on NetCDF and data format conventions.

- Interactions w/ CSU collaborators on refinement of visualizations and analysis needs.
- Installation of plugin. Confirm that CSU/PNNL researchers can analyze data on `davinci.nersc.gov` as well as on their local workstations.
- Incorporate Morton-ordering (space-filling curve) layout for efficient file access and multi-resolution rendering.

2.7 Climate: 3D Visual Data Analysis in ESG's Climate Data Toolkit

VACET is committed to supporting the needs of the Community Climate System Model (CCSM) Consortium in collaboration with the Earth System Grid.

Mission. The most advanced climate modeling systems seek to enable a new deeper understanding of the dynamics of global carbon cycle, atmospheric chemistry, land and ocean ecological processes and their coupling with climate. This will allow pursuing reliable answers to fundamental questions related to climate variability and global change at time scales ranging from decades to centuries. In this effort VACET will work in close collaboration with the Earth System Grid and provide new advanced data analysis and visualization tools to the CCSM Consortium and the climate modeling community in general. One target will be the deployment of a first set of tools in FY08, in time to facilitate the analysis of data for the 5th assessment report of the Intergovernmental Panel on Climate Change (IPCC).

Team

- Valerio Pascucci (Utah), Team Leader.
- Peer-Timo Bremer (LLNL): core visualization techniques, data comparison, topology.
- Marty Cole (Utah) deployment of visualization libraries.
- Jamison Daniel (ORNL) user interaction, case studies, testing of tools.
- Ming Jiang (LLNL) topological analysis, VTK components.
- Daniel Laney (LLNL) data management and streaming.
- Ajith Mascarenhas (LLNL) feature tracking.
- Claudio Silva (Utah) VisTrails.
- Xavier Tricoche (Purdue) 2D topological visualization and analysis.

Stakeholder Needs

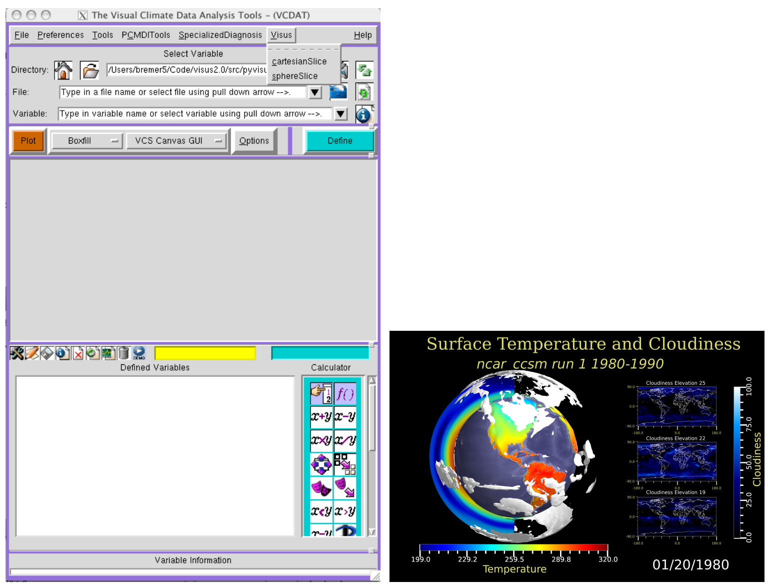
1. Deploying advanced visualization capabilities into the CDAT tool and create a clear path for similar integration in other tools.
2. Extend the visualization software to incorporate domain specific requirements, data formats, and vector field visualization.
3. Support time-dependent and cross-dataset comparison, visualization and analysis.
4. Develop new analytic capabilities for climate data (first deployed into CDAT/VCDAT).

5. Integrate with VisTrails advanced framework for tracking and logging of internal state and provenance information.

6. Develop a visualization and data analysis scenario for understanding of complex coupled phenomena such as the multi-scale dynamics the complete carbon cycle on earth.

Accomplishments

1. The ViSUS 2.0 framework has been released as part of CDAT 5.0. This result follows from a substantial amount of software engineering effort over the past 18 months. This result is significant because all CDAT users now have access to a solid set of fundamental 3D visualization capabilities.



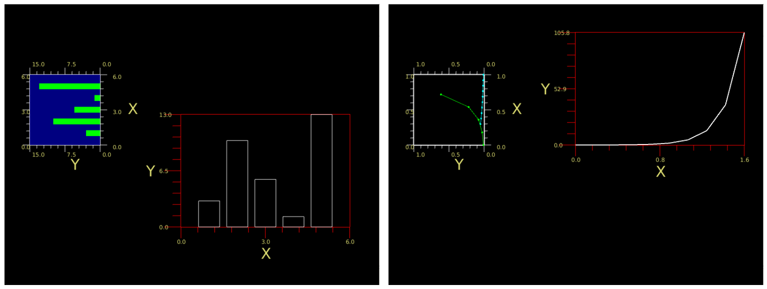
(a) Custom VCDAT interface showing the new ViSUS capabilities.

(b) An example scene showing 3D isocontouring, slicing, and multiple 2D plots.

Figure 16: New 3D visualization capabilities are now part of the production CDAT release.

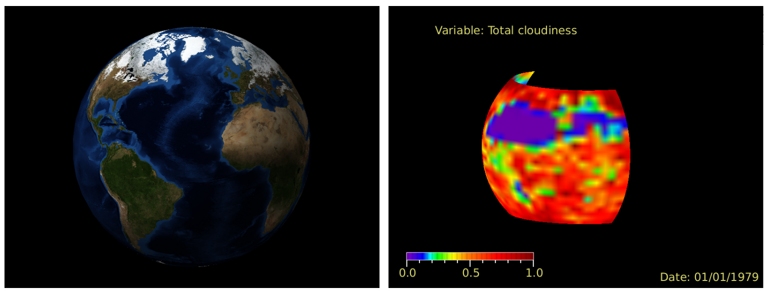
2. We have further expanded the ViSUS code base to include more advanced features:

   - In response to a major request from stakeholders, we have replaced the FOX GUI toolkit with a light weight FLTK 2.0 based system.
   - Added system dependent off-screen rendering capabilities.
   - Added support for histogram and graph drawings (see Figure 17).
   - Added support for multi-modal height field rendering.
   - Added support for zoom-in views.
   - Added complete and partial sphere mappings.
   - Added Polyline rendering.
   - Added support for general mesh display.



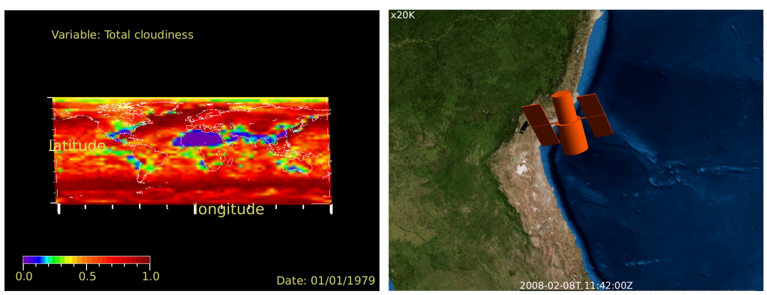
(a) Histogram display. (b) 1D Function display.

Figure 17: At the request of the science stakeholders, we have added the ability of the new visualization infrastructure to support commonly used operations (histogram and graph plots).



(a) Earth display using satellite imaging. (b) Partial mapping of cloudiness.

Figure 18: New features include complete (left) and partial (right) sphere mappings.



(a) Continents rendered using the new polyline display capability. (b) A satellite above the earth model rendered from a triangular mesh.

Figure 19: New features include polyline rendering (left) and support for arbitrary triangle meshes (right).

- Added example scenes and documentation.
- Integrated ViSUS into CDAT's build system.

Future Work

Near Term

- Extend functionality as requested by the climate community.
- Re-integrate volume rendering into ViSUS 2.0.

- Develop climate specific user interface.
- Provide example scripts as tutorials for new users.

Medium Term

- John Drake and Phil Jones have requested general z-mappings to support data simulated on constant air pressure planes rather than constant elevation plane.
- John Drake and Phil Jones have requested support for mapped logical grids to be supported by ViSUS.
- Extend ViSUS library to natively support time-dependent data.

Longer Term

- Provide library to dump simulation output directly into IDX format.
- Provide command line interfaces to convert data into IDX format to allow streaming visualization.
- Remote visualization client.

2.8 Climate: Improving Workflow Efficiency in ESG's CDAT

Team Members

- VACET: Claudio Silva, Emanuele Santos, Lauro Lins, and others (Utah).
- ESG: Dean Williams (LLNL).

Approach

Work on the CDAT and VisTrails integration started around November 2007. After discussions with Dean Williams (leader of the CDAT team), we decided to integrate the CDAT tools with VisTrails as a VisTrails package. The idea is to allow the CDAT users to take advantage of the provenance mechanism as well as the visual programming interface to build CDAT workflows. The VisTrails CDAT package will contain modules for each of the CDAT subsystems (cdms, cdutil, vcs, etc.).

Because of the complexity of these subsystems, it is impractical to wrap all of the functions manually. In order to solve this, the CDAT team has been modifying CDAT to generate XML descriptions for all its subsystems, with the idea that VisTrails will parse them to generate the modules necessary to build the workflows.

Accomplishments

Currently, the vcs (Visualization and Control System) subsystem is generating XML descriptions for its functions. The VisTrails team has written the parser, and vcs is automatically wrapped in VisTrails. To test the framework, we also manually created a minimal set of XML descriptions for the cdms subsystem so we could build a complete CDAT workflow in VisTrails.

At this point, users can create workflows and visualize the results either on the standard CDAT Window or in the VisTrails Spreadsheet as a static image. Our goal is to embed the CDAT window as a spreadsheet cell widget so users can directly interact with it. One of the technical difficulties

in this case is that CDAT does not run natively on some platforms, instead, it relies on the X Windows and Tcl/TK while VisTrails runs native on each supported platform since it is built on top of Qt.

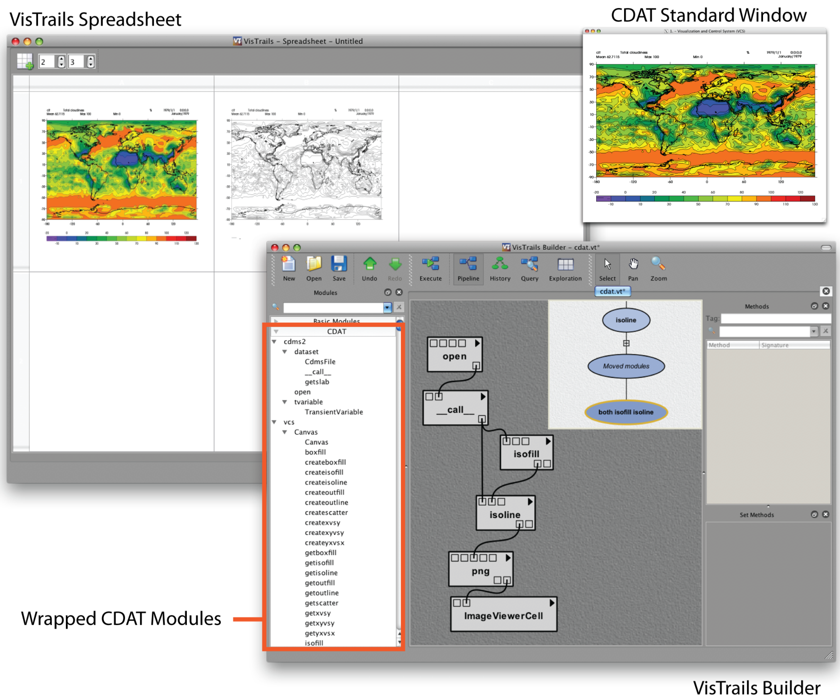


Figure 20: Example of a CDAT workflow built in VisTrails. In the VisTrails builder it is shown a workflow (and its history tree) for generating a simple overlay plot using CDAT's isofill and isoline graphics methods. Notice the list of CDAT functions wrapped as VisTrails modules on the left. The results can be shown in the VisTrails Spreadsheet or in CDAT's standard window.

2.9 Fusion: Particle and Magnetic Field Visualization and Analysis

VACET is engaging with several fusion stakeholders in an effort that includes a recent SAP project (led by VACET's Allen Sanderson (Utah)).

Particle Path Analysis and Visualization. The physicists are currently generating simulations that use millions to billions of particles, with each particle containing multiple scalar and vector data (multivariate data). They would like to have the ability to explore the nature of the particle orbits in an interactive manner. It is impractical to view billions, much less millions, of particles at one time and glean any insight. As such, the physicists would like to have tools that allow them cull the particles using a user defined query or other statistical tools.

At the same time as the particles are displayed, physicists are interested in seeing the particles in context of data that is not associated with the particle but is part of the simulation. This data may be scalar (electric potentials) or vector data (magnetic field) and may have its own visualization requirements. For instance, the scalar data may be viewed using a variety of techniques from volume rendering to slicing.

Magnetic Field Analysis and Visualization (in collaboration with the Fusion SAP). Physicists are currently studying the affects of magnetic islands that form in the plasma. These islands cause defects in the magnetic field and the current flow resulting in contact between previ-

ously separate regions. This contact results in “hot areas” coming into contact with “cool areas,” which leads to core cooling. Physicists would like to have tools that allow them be able to automatically generate Poincar maps of the magnetic field and detect the island formation and track them over time.

At the same time as the Poincaré maps are displayed, physicists are interested in seeing them in context of data that may or may not be associated with it, but is part of the simulation. This data may be scalar (electric potentials) or vector data (magnetic fieldlines) and may have its own visualization requirements. For instance, the scalar data may be viewed using a variety of techniques from volume rendering to slicing.

Comparative Analysis and Visualization (in collaboration with the Fusion SAP). As physicists develop and refine their simulation codes, they need tools for performing intra- and inter-simulation comparative analysis, as well as comparison between simulations and experiments. These tools will need to be able to analyze scalar and vector data produced on different meshes and different time scales. At the same time physicists desire tools that will allow them to compare and visualize multivariate data.

The Team

VACET

- Allen Sanderson (Utah), Team Leader.
- Sean Ahern and Dave Pugmire (ORNL).
- Prabhat and Gunther Weber (LBNL).

Science Stakeholders - Particle Path Visualization and Analysis

- Stephane Ethier (PPPL) of the Micro Turbulence SciDAC (Pat Diamond - PI).
- Seung-Hoe Ku (NYU) and Julian Cummings (CalTech) of the Edge FSP SciDAC (C.S. Chang - PI).

Magnetic Field Analysis (work in partnership with the Fusion SAP).

- Scott Kruger (Tech-X) and Josh Breslau (PPPL) of the CEMM SciDAC (Jardin - PI).
- Bill Nevins (LLNL) of the Micro Turbulence SciDAC (Bill Nevins - PI).
- Don Bachelor (ORNL) of the CSWIM SciDAC (Don Bachelor - PI).

Comparative Visualization and Analysis (work in partnership with the Fusion SAP).

- Steve Jardin and Josh Breslau (PPPL) of the CEMM SciDAC (Jardin - PI).

Accomplishments

The major emphasis in this period has been on shifting the visualization work over to VisIt from SCIRun. This work has included:

- Leveraging the query based mechanisms (parallel coordinates) developed for accelerated particles (Section 2.1) for use with fusion particles.

- Identification of a data model strategy for indexing and organizing particle data into H5Part files.
- Working VisIt prototype that can:
  - Load H5Part indexed data into VisIt.
  - Interactively execute efficient queries.
  - Display/Refine parallel coordinates (see Figure 21).

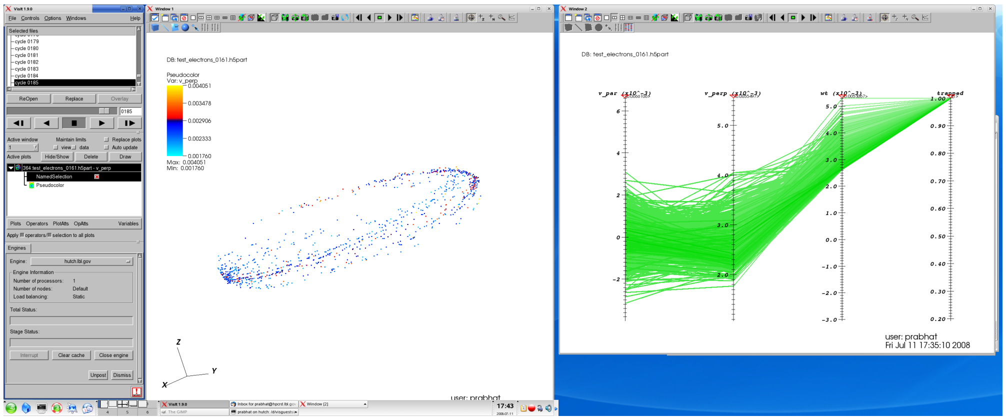


Figure 21: Screen snap shot of query based fusion particle visualization using parallel coordinates.

- Other items:
  - Visit to PPPL to give a VisIt tutorial – PPPL no longer has an in house visualization expert and as such requested a tutorial to help get their researchers up to speed in using VisIt. Approximately 20 physicists and staff attending the tutorial.
  - As part of our on going outreach, visits were made to FACETS all hand meeting as well as to PPPL to discuss physicist's needs.

2.10 Fusion: Community-wide, Production Quality Visual Data Analysis Software

Tech-X has identified a need for production quality visualization for fusion and accelerator simulation modeling delivered within VisIt. They have several different applications and have expressed the desire to have one fully-featured tool that meets their needs for fusion and accelerator modeling projects. Due to the combination of features and performance capabilities, as well as the presence of VACET as a vehicle for developing new capabilities and providing support, they would like VisIt to be the tool of choice for their work.

A list[6] of milestones and deliverables was generated by Sean Ahern during a visit to Tech-X in May of 2007. Dave Pugmire visited Tech-X in January of 2008 to prioritize and clarify their needs. Dave also participated in their all hands meeting in September of 2008. In addition, Tech-X frequently makes dynamic requests, often from FACETS PI John Cary, who actively monitors the visit-developers and visit-users mailing lists.

[6] `http://www.sci.utah.edu/vacetwiki/index.php/Collab:Fall07_VisIt_Fusion_Visualization`

The Team

VACET Team

- ORNL: Sean Ahern (team leader), Dave Pugmire, Jeremy Meredith.
- LLNL: Hank Childs, Brad Whitlock.
- LBNL: Gunther Weber.
- UC Davis: Christoph Garth.
- Utah: Allen Sanderson.

Stakeholder: Researchers at Tech-X Corporation.

Accomplishments

Poincaré Analysis Tools
Poincaré analysis is a tool that can be used to aid in the understanding of magnetic field line flows, field line winding and island identification and classification. A stand alone library has been developed by Allen Sanderson that provides this functionality and integration of this into VisIt is desired. Incorporating this into VisIt is staged in several separate steps. We have focused on the streamline engine (see Section 3.1) for streamlines as well as the engine for generating Poincaré plots.

Better Streamlines

- (A1.1 Streamline direction (forward, backward, both) must be user controlled. See Figure 22.

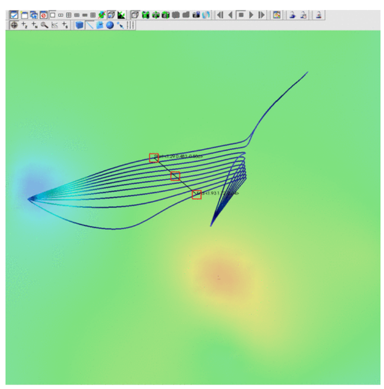


Figure 22: Multi-directional streamlines.

- (A1.2) Termination criteria. Streamlines must be generated to arbitrary lengths.
- (A1.3) Rich set of integrators. Currently, VisIt supports 5th order Runge-Kutta and Adams-Bashforth multi-step integration. Other methods, suitable to specific problem domains that can be selected by the user are needed. This task is work in progress.
- (A1.4) Parallel streamline algorithms. To scale to problem sized data, parallel streamline algorithms need to be evaluated and implemented. This task is work in progress.

Customized Interfaces FACETS PI John Cary contends that the VisIt interface is too complicated for his user community. Although he values VisIt's visualization capabilities, he wants a streamlined interface that is similar to his previous visualization tool.

- (F1) Prototype a streamlined interface.
- (F1.1) Re-architect portions of VisIt to allow a separate program to contain the OpenGL context.
- (F1.2) Write a prototype application that utilizes this new design. See Figure 23.

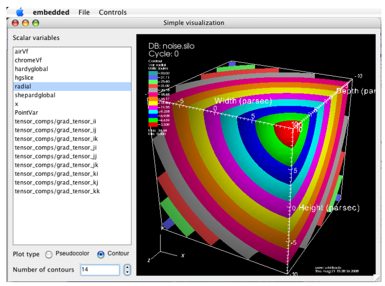


Figure 23: This shows the prototype application written for Tech-X (F1.1 and F1.2). A custom interface allows for users to choose variables and plot types. VisIt's process that normally does rendering now renders into the context of this custom skin interface. This prototype was implemented by Brad Whitlock. David Pugmire demonstrated this prototype at the FACETS AHM in September 2008 and the amount of interface customizability now possible was very well received.

Future Goals

We expect to deliver A2 (Poincare plots). If we decide to pursue F2, then that would also happen in the next 6 months. H1 should also happen within the next 6 months as NERSC Analytics work for Cameron Geddes depends on it.

We expect more "1000 grains of sand" (numerous small feature requests) deliverables to be added to section G that will need to be turned around in the next 6 months.

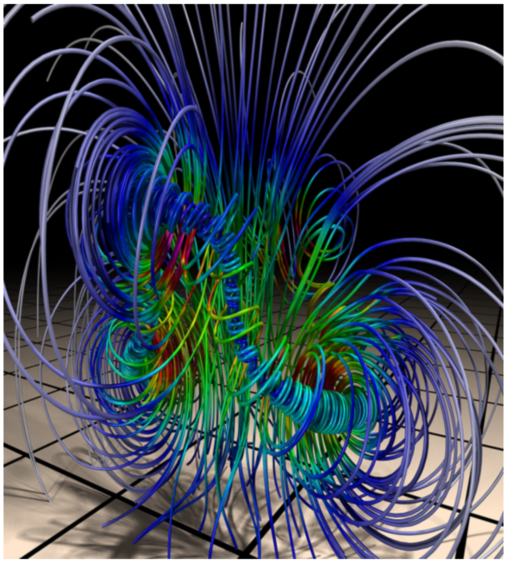
3 Common Infrastructure Projects

3.1 Streamlines

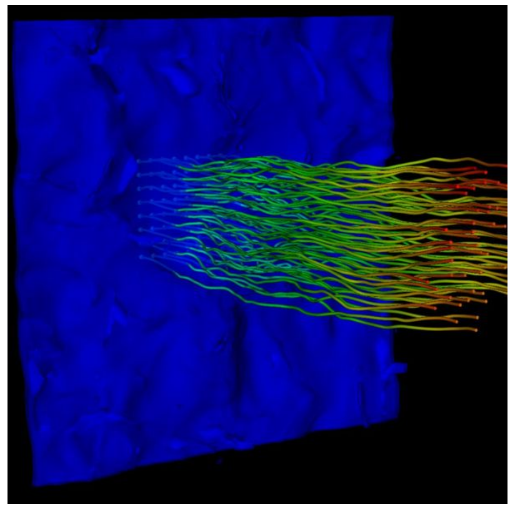
A long-standing gap in visualization technology in general, and the VACET technical portfolio in particular, is the computation of streamlines in parallel and on multi-grid domains. Recent work has produced a production-quality, parallel capable "streamlines engine" that meets multiple VACET science stakeholder needs as well as makes a novel contribution to the field of high performance visualization. The new parallel-capable, multi-grid aware streamlines engine has been deployed in VisIt (version 1.10), VACET's production-quality parallel-capable visual data analysis software infrastructure.

Science Stakeholders

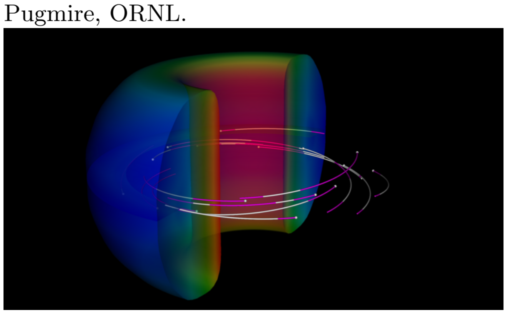
VACET is working with the following stakeholders, all of whom have requested and stand to benefit from parallel and multi-grid streamlines: (1) Phil Colella (LBNL) of the SciDAC Applied



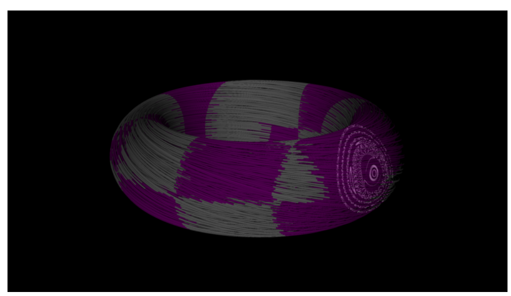
(a) Two merging vortex cores computed by the APDEC/Chombo code. VisIt computes the streamlines from the AMR dataset, then we render them with an off-the-shelf raytracing (photorealistic) software application. Dave Pugmire, ORNL.



(b) Image produced by Dave Pugmire and Sean Ahern (ORNL) for the SciDAC "Shocks" Center.



(c) Seeds placed randomly within the tokamak are used for streamline generation.



(d) Poincaré plot of randomly seeded streamlines in data produced by a fusion tokamak simulation.

Figure 24: These images show application of the VACET streamlines engine to data from different science domains: combustion (upper left), shock physics (upper right), and fusion (lower left and right).

Partial Differential Equations Center (APDEC); (2) John Cary (Tech-X) of the SciDAC Science Application "Framework Application for Core-Edge Transport Simulations" (FACETS); (3) Steve Jardin and Stephane Ethier (PPPL) of the SciDAC Center for Extended Magnetohydrodynamic Modeling (CEMM); (4) Don Batchelor (ORNL) of the SciDAC Science Application "Simulation of Wave Interactions with Magnetohydrodynamics" (SWIM); Johan Larsson and Sanjiva K. Lele (Stanford) of the SciDAC Science Application "Simulations of Turbulent Flows with Strong Shocks and Density Variations." Although we have had direct contact and requests with these specific SciDAC projects, virtually all our stakeholders will stand to benefit from this technology.

Science Stakeholder Needs.

1. AMR/multi-grid domains. Streamlines computation consists of integrating a path tangent to a vector field. Traditionally, these algorithms assume a single domain, and they do not accommodate the boundary conditions that occur at coarse/fine grid boundaries in AMR datasets. These boundary conditions impose special challenges: (1) streamline algorithms do not understand nesting of patches, and do not accommodate the transition from finer patches from coarser patches; (2) the resulting streamlines do not have the desired level of continuity across grid boundaries; (3) there is little information as to streamline quality in general and in particular when streamlines cross patches.

2. Poincaré plots. This type of visualization (see Figure 24) is useful in fusion to quickly visually identify magnetic "islands" in the plasma core.

3. Parallel Performance. The long-term objective is to leverage parallel computing platforms to accelerate streamlines computation and to accommodate ever-large datasets.

4. Pathlines. Accurate streamline computations across timesteps.

5. Support for additional/custom Initial Value Problem (IVP) solver types with advanced features (e.g. continuous output), etc.

Accomplishments this Period.

1. Introduce new IVP solver class hierarchy into VisIt with the aim of implementing individual IVP schemes as subclasses.

2. Incorporate several integration schemes such as RKF45 and DOPRI56 (and possibly tie in IVP libraries such as CVODE).

3. Introduce streamline (convenience) class more specifically aimed at streamline integration requirements (i.e. construction and storage of a global IVP solution) and a matching datatype to pass streamline sets between VisIt filters and plugins.

4. Parallelize IVP solver and streamlines.

5. Implement and evaluate parallel streamline algorithms.

6. Adapt streamline integration to multiblock datasets.

7. Deploy in VisIt version 1.10 (See Figure 24).

Future Goals.

1. Parallel Performance. While we have an initial prototype in production, we would like to better understand and characterize its performance. Since there exists no prior work in this area, we expect significant results: one or more major publications, benefit to our science stakeholders.

2. AMR streamlines. While our initial implementation is multi-block aware, a good deal of work remains: (1) continuity at and across coarse/fine grid boundaries; (2) comparison of different interpolation basis functions; (3) evaluate alternatives using objective error metrics.

3. Apply the streamlines engine to science problems (e.g., Poincaré analysis).

3.2 High Quality Volume Rendering

This ongoing project aims to perform "technology transfer," where we perform the software engineering necessary to transition a research prototype code into a form suitable for production deployment in VisIt, our production visualization application. This project, which involves VACET personnel from Utah, LLNL, and ORNL, is a successful example of focusing software engineering effort to bring a successful research prototype into production use. The technology in question is SLIVR: the SCIRun Library for Interactive Volume Rendering[7]

Accomplishments

- SLIVR is now completely integrated into VisIt. The additions currently reside in the development trunk of VisIt, and are expected to ship with the upcoming 1.11 VisIt release. Examples images show SLIVR applied to 3D medical data (Figure 25) and the GUI for multi-dimensional transfer functions (Figure 26).

4 Technology Incubation Projects

4.1 Query-Driven Visualization of Time-Varying AMR Datasets

As part of our research portfolio, we are exploring alternative techniques for extending the fundamental idea of Query-Driven Visualization to single- and multiple-timestep Adaptive Mesh Refinement (AMR) data. Several of our stakeholders generate AMR data (e.g., APDEC) and have expressed interest in leveraging emerging high performance query-driven visualization technology to accelerate their scientific knowledge discovery.

Recent research has produced a new approach that enables query-driven analysis and multitemporal visualization of time-varying AMR data. Previously, such analysis and visualization efforts were hindered by the dynamic temporal and spatial properties of AMR grid hierarchies. We developed a two-step method for compositing and synchronizing AMR data from a series of timesteps.

We first generate a composite template from the AMR grid hierarchies of these timesteps; the composite template preserves the finest level of grid cell refinement from each grid hierarchy. We then synchronize each timestep's grid hierarchy to the composite template. This approach enables our method to process queries on a common AMR grid hierarchy. Using this data structure, we

[7]See `http://slivr.sci.utah.edu`.

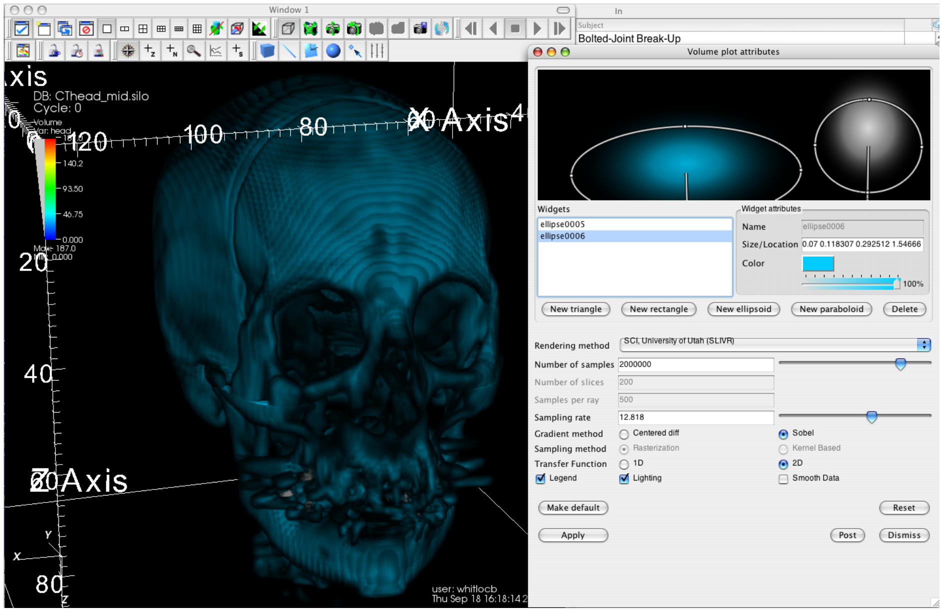


Figure 25: SLIVR is now integrated in VisIt and appears as one of its volume rendering options.

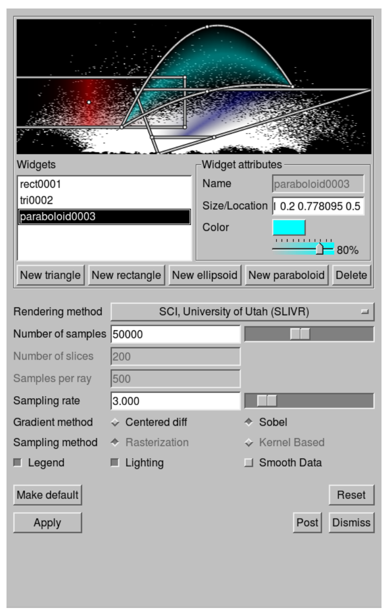


Figure 26: SLIVR provides a scalar magnitude/gradient 2D histogram and a variety of widgets to control rendering operation.

move the work of query processing to the GPU to realize the benefit of greatly accelerated QDV analysis. On the GPU side, we integrate our new method with a GPU-based query engine, called the Bin-Hash index (see Section 4.2).

Our method facilitates query-driven analysis of time-varying AMR data, and generates two types of time-dependent visualization: temporally sequential and temporally concurrent. In temporally sequential visualizations, features from each timestep are analyzed and visualized individually in

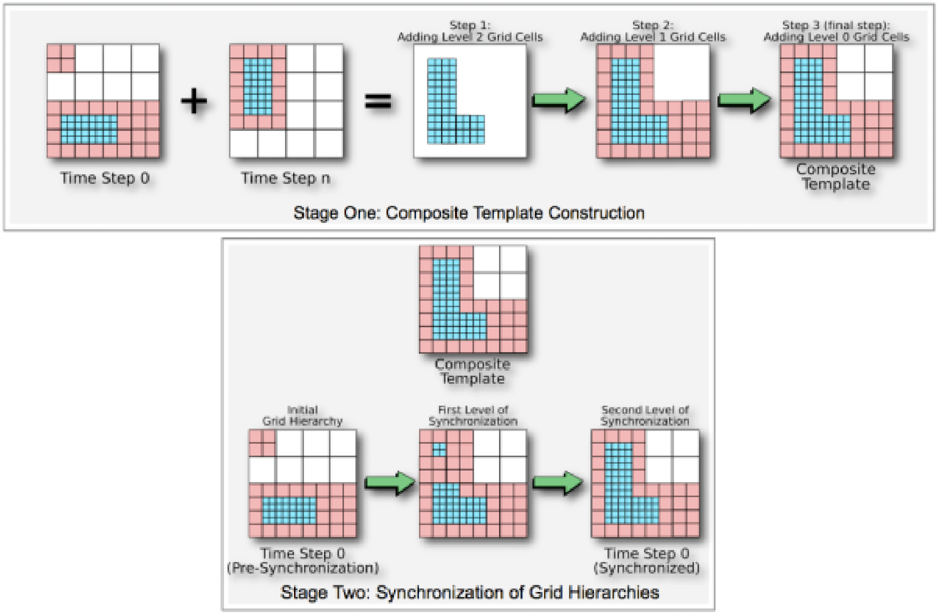


Figure 27: Performing QDV on time-varying AMR data requires using a new data structure that reconciles potential/likely topological discrepencies between different AMR grids.

sequential frames as an animation. Comparatively, in temporally concurrent visualizations, a single multitemporal image conveys how queries characterizing important features evolve over time. In temporally sequential visualizations, our GPU-based QDV engine enables accelerated analysis; users can process queries over multiple time steps and view the results in real-time as an animation.

The primary contributions of this work are:

1. We develop a new framework for doing QDV processing and visualization of time-varying AMR data. The core of this method is based upon a synchronization strategy that addresses the disparities in spatial refinement that exist between any series of timesteps in an AMR-based simulation.

2. We demonstrate the first GPU-based QDV approach that utilizes a GPU-based indexing strategy to accelerate query processing, efficiently utilize GPU memory, and accelerate QDV methods.

Results

The following images in Figure 27 depict our two step process that facilitates query-driven visualization of time-varying AMR data.

The figure at left illustrates the sequential process of compositing the AMR grid hierarchies of two selected timesteps. The process begins by filling the composite template with all grid cells, from both timesteps, of the finest level of refinement. In each subsequent pass, our procedure adds grid cells of the next level of lesser refinement to the template - conditioned on the basis that a more finely refined grid cell has not already been placed at that position. Finally, we add grid cells of the coarsest level of refinement to the template.

The figure at right depicts the sequential process of synchronizing the grid hierarchy of a given timestep with a composite template. At each level of synchronization, grid cells conditionally

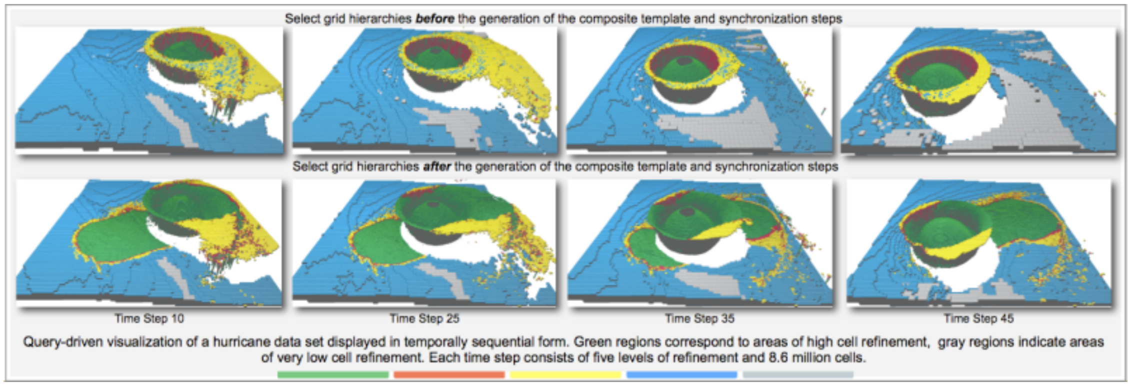


Figure 28: This series of images, selected from 48 timesteps, compares query results from non-sychronized (top row), and sychronized (bottom row) AMR grids of the Hurricane Isabel dataset. The query used on each timestep consists of two parts; we query for regions of low pressure ($-200 \leq pressure \leq 20$) OR regions of high pressure ($500 \leq pressure \leq 1000$).

refine themselves by one additional level according to whether or not they are synchronized with the composite template. In this example, synchronization is complete for the grid hierarchy in the second level of synchronization.

Publications

Luke Gosink, John Anderson, E. Wes Bethel, Kenneth I. Joy, Query-Driven Visualization of Time-Varying Adaptive Mesh Refinement Data, in: IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization), 2008.

4.2 Bin-Hash Indexing

To support high-performance query-driven visualization, we have undertaken research aimed at exploring fundamental algorithms and data structures that will allow us to leverage emerging computing architectures to achieve very high levels of performance. This project has produced a novel approach for encoding/searching data, called the “Bin-Hash Index,” that runs at very high levels of performance on commodity GPUs.

Though GPUs offer tremendous parallelism, their utility for database tasks is limited by a small store of resident memory. For example, the largest amount of memory available on NVIDIA Existing GPU-based works that utilize the projection index to answer a query are thus significantly limited by GPU memory resources. Our Bin-Hash index presents one method for ameliorating the challenges imposed by limited GPU memory. The Bin-Hash index uses a form of compression, implemented through a multi-resolution representation of the base data information. This compression strategy allows us to query dataset sizes that would otherwise not fit into the memory footprint of a GPU were we to use a traditional projection index strategy.

In the Bin-Hash approach, we bin the base data for each column and generate a spatial hash of column values in each bin. To resolve a query condition on a column, we first access the bin numbers. For range conditions such as “pressure $>$ 100”, we determine which bins satisfy the condition and which ones don’t, based on the boundaries of the bins. There is only one exception: the bin containing the value 100. We call such bins “boundary bins.” We need to examine the colum values for the records in the boundary bins to determine whether each actually satisfies

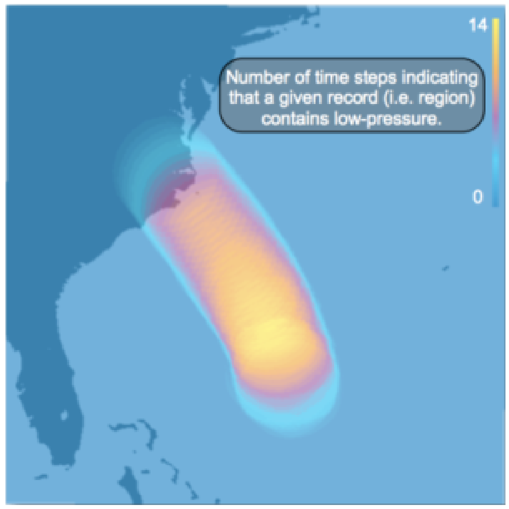


Figure 29: This multitemporal image depicts summary statistic information gathered from queries processed over 48 timesteps of the Hurricane Isabel WRF Model dataset. In this multitemporal image, hurricane direction and velocity information are conveyed by querying for regions of low-pressure.

the query condition. We call the records in the boundary bin the candidates and the process of examining the candidate values "the candidate check." Altogether, to answer a query, we access the bin numbers and the base data values of the records in boundary bins. The data contained in the summed total of both these data structures is much smaller than the column projections used by other strategies that employ the GPU to answer a query. Additionally, the procedure of examining the bin numbers and the process of performing the candidate checks offers the same level of concurrency as the projection index, and achieves excellent performance as we demonstrate later.

The main contributions of our work are the following.

- We introduce the Bin-Hash data structure for accelerating selection queries using GPUs; existing work on processing such queries only uses the projection index, which requires using much more (GPU) memory than the Bin-Hash approach.
- In our performance tests, our approach is shown to outperform the fastest indexing strategy used for our key application, Query-Driven Visualization ; earlier works lack such a direct comparison.
- We demonstrate the utility of the perfect spatial hash[8] as a parallel data structure. In our tests, thousands of threads concurrently and efficiently access partitioned base data on a parallel processor. Additionally, we show how this spatial hashing data structure is essential for our Bin-Hash index to reduce the amount of data needed on the GPU, and to utilize the GPU's parallel processing capability.

Results

The results, shown in Figures 30, 31, and 32, indicate that our approach shows a great deal of promise for accelerating queries on a GPU. It outperforms FastBit, and because of the compact

[8]Ed: The Perfect Spatial Hash is an algorithm and data structure described by H. Hoppe in a 2006 Siggraph paper.

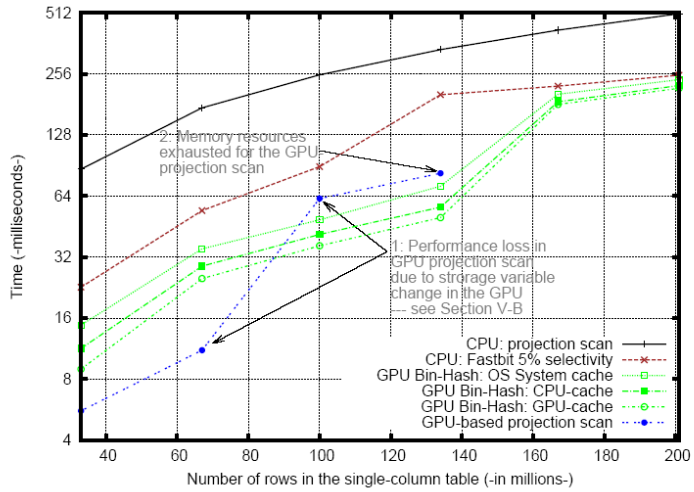


Figure 30: This figure shows the timing results of 3 different querying strategies. The x-axis depicts the number of rows contained in each single-column table queried. Each table is queried with 5 separate queries selecting 1%, 5%, 10%, 20%, and 40% of the database records as hits. The projection scan and the Bin-Hash methods displayed no difference in timings with respect to increasing query selectivity. This is due to the fact that all of these strategies have a working complexity of O(n). For comparison, the FastBit results for queries returning 5% of the records as hits are shown in both this table, and in Figure 31. For clarity, the performance lines for the three cache levels of the Bin-Hash index are intentionally colored the same to reflect the range of performance of the Bin-Hash index.

nature of the underlying data structure, accommodates much larger source data than is possible with a projection scan on a GPU.

Collaborations

This project is a collaborative involving:

- VACET Researchers at UC Davis and LBNL.
- SciDAC Scientific Data Management Center: Kesheng Wu (LBNL).
- SciDAC Institute for Ultrascale Visualization: John Owens (UCD).

4.3 Embedded Boundary/Material Interface

In many applications it is necessary to reconstruct or track the boundary surfaces, or "interfaces," between multiple materials that commonly result from multi-fluid Lagrangian-Eulerian hydrodynamics calculations. This problem, where the generated data sets have the characteristic that each cell contains a "fraction" representing the percentage of each material contained in the cell, now arises in a variety of applications, and is frequently called the **embedded boundary problem**. The challenge is to utilize the material fractions in each cell to reconstruct the boundaries between materials. VACET researchers have been developing a number of possible solutions to this problem.

Having a production-quality solution to this problem will allow one of our primary stakeholders and his team (Colella/APDEC) to completely adopt VisIt for their visual data analysis infrastruc-

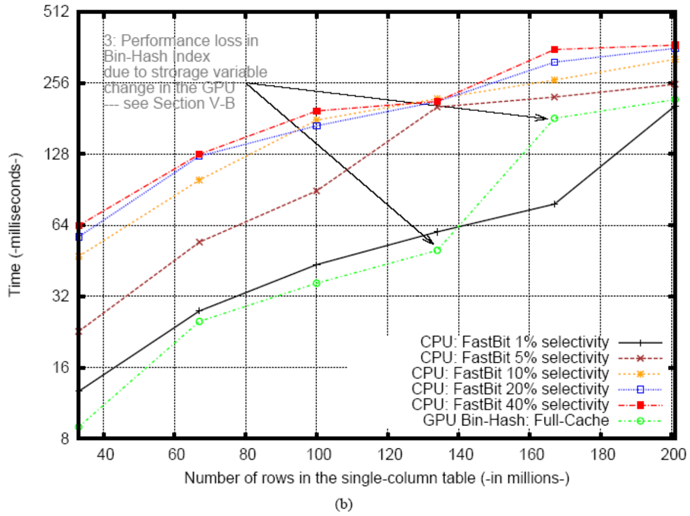


Figure 31: This figure shows the timing results of FastBit, shown here when utilized by 5 separate queries, respectively selecting 1%, 5%, 10%, 20%, and 40% of the records as hits. We observe that FastBit shows significant timing differences with respect to the selectivity of the query. For comparative purposes with Figure 30, the Full-Cache results for the Bin-Hash method are shown. It can be seen that with queries selecting 1% of the records as hits, FastBit performs remarkably well to the parallel approach equaling or even besting the best case of the Bin Hash strategy when querying some of the largest tables examined.

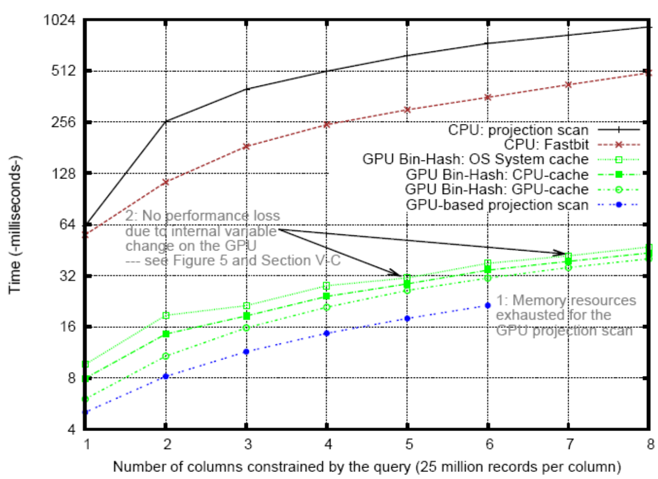
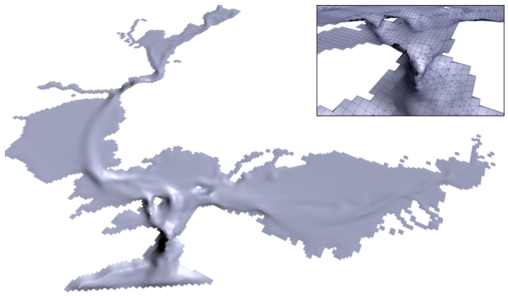
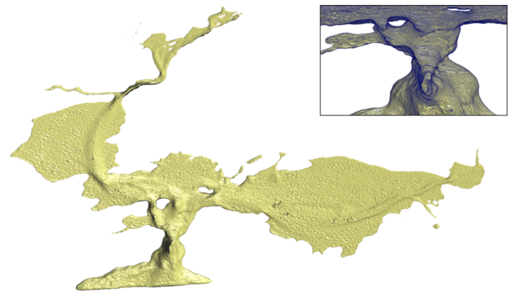


Figure 32: This figure shows the results of 3 different querying strategies over a table containing 48 columns and 25 million rows. The X-axis of this table (1, 2 . . . 8) lists the total number of columns constrained in the multidimensional query. The selectivity of these queries grows at constant rate: each column added to the query selects an additional 12% of the rows from the table being queried.



(a) San Francisco Bay bottom using legacy VisIt implementation of material interfaces.



(b) San Francisco Bay bottom using new material interface methods.

Figure 33: Comparison of legacy (left) and new (right) techniques for computing embedded boundaries.

ture. The benefit to their team is a substantial cost savings – they will no longer have to create, update and maintain their own code base for visual data analysis.

We expect to develop a number of solutions to this problem. We do not expect that any one algorithm will complete solve this problem, but that we must produce a number of solutions that can impact our stakeholders. There are several problems that must be solved: (1) the interface reconstruction method must produce interfaces that preserve the given volume fractions, (2) the solution must be scalable, (3) the solution must be computationally "quick," and (4) adaptive techniques must be developed that can be tailored to the user's requirements. We will implement this into VisIt to get the maximum impact with the scientists.

Accomplishments this Period

1. We have developed a new adaptive that utilizes an active interface method to adjust an initial approximation of the interface to one that matches given volume fractions (see Figure 33. This allows us to separate the problem into two components: a topology-generator that develops a plausible topology for the interface reconstructions – from this we can generate an approximate interface – and an iterative system that adjusts this initial approximation to match the given volume fractions.

2. A paper has been submitted on this new algorithm as John Anderson, Christoph Garth, Mark A. Duchaineau and Kenneth I. Joy, "Smooth, Volume-Accurate Material Interface Reconstruction," submitted to IEEE Transactions on Visualization and Computer Graphics, September 2008.

3. We have developed an initial approximation method based on a "Potts" Model. This allows us to generate interfaces that reflect some fine detail in the scene.

4. The Potts model topology generator has been published as John Anderson, Christoph Garth, Mark A. Duchaineau and Kenneth I. Joy, "Discrete Multi-Material Interface Reconstruction for Volume Fraction Data," Proceedings of the European Visualization Conference, Computer Graphics Forum, Vol. 27, No. 3, 2008, 1015-1022.

5. We have implemented earlier material interface codes (e.g., PLIC) in VisIt for our APDEC Stakeholders.

Future Work

1. Implement the new Active Interface method into VisIt.
2. Develop an AMR version of the Active Interface method that can be used by our APDEC collaborators.
3. Develop a multiresolution version of the Active Interface algorithm.
4. Longer term: fully implement these algorithm into VisIt, develop a method for unstructured meshes (the topology generator is the difficult component here), and develop the level-set method. Each of these project have been started, and some preliminary results have been obtained. However, we expect this project, and its "spin-offs to be active during the entire VACET project.

4.4 Uncertainty Visualization

Towards the Visualization of Multidimensional Probabilistic Distribution Data

Uncertainty information is an important characteristic associated with much of the data scientists encounter. Such information is typically included with the data as 2D charts and graphs, however incorporating uncertainty into visualization techniques has proved quite challenging. Recent work examines a class of uncertainty data that is characterized as a set of probability density functions (PDFs) defined across a triangular mesh, and explores ways at visually presenting this data.

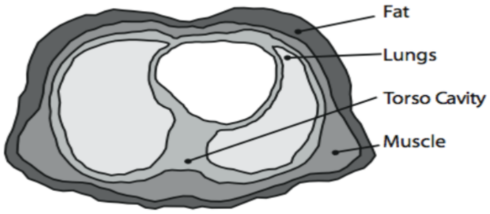
The data used in this work comes from the sensitivity analysis electrical conductivity within computational models of bioelectric fields of the heart. Such an approach quantifies the sensitivity of the electrocardiographic forward problem by creating a mathematical model to reconstruct a biological experiment in which the voltages on the human torso are estimated based on the input electrical conductivities (from previous work). The simulation stochastically varies the input conductivities of different tissues such as fat, lungs, or muscle and examines the resulting changes in potential across the torso.

Our recent visualizations strive to investigate the complex sensitivity analysis data. The data is defined on the classified torso mesh (shown below in Figure 34). The mean and standard deviation of the data are shown colormapped onto the torso data space, and the first approach we explore is to combine the mean and standard deviation into a single visualization. We achieve this by simply encoding the mean into a heightmap, and colormapping variance onto this map. This provides for a simple way to show two variables of the dataset, and quickly find locations of large and small average values of the data, as well as high variance values.

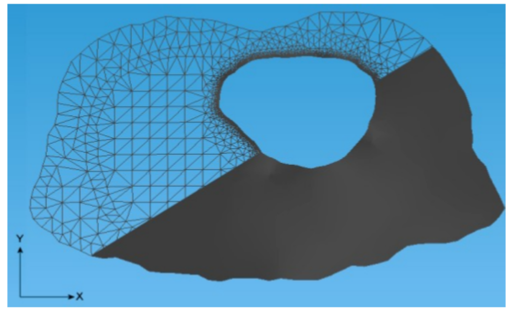
The next approach we investigate is to create a volume by stacking the output potentials for every input conductivity. This volume can be volume rendered, or iso-surfaced. The structure of the iso-surface is of most interest. An iso-surface that falls straight down indicates the potential at this point in 2D space does not vary when changing; thus, it is considered independent at that point. A bending iso-surface indicates areas of high dependence on the input conductivity. We can also see that even in the region with the highest variance only small conductivities result in potential changes. This relationship becomes visible using our new visualization technique (Figure 35).

Visualizing Summary Statistics and Uncertainty

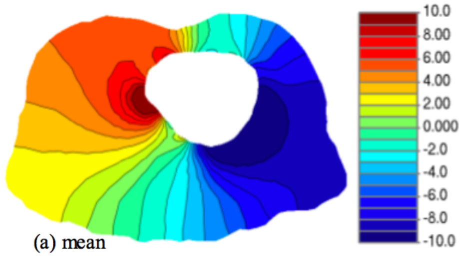
An important visualization research problem is to effectively convey uncertainty information along with traditional visual data representations. Recent work investigates the problem from a graphical



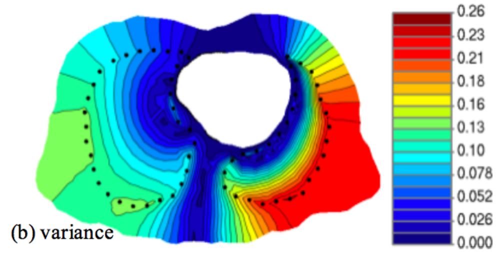
(a) The classified human torso.



(b) The torso mesh space.



(c) Mean colormapped onto the torso.



(d) Standard deviation colormapped onto the torso.

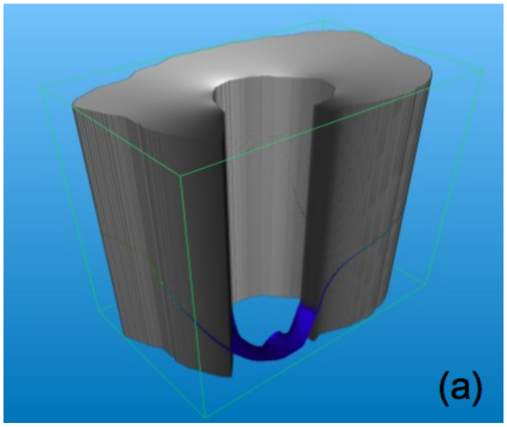
Figure 34: Comparison of legacy (left) and new (right) techniques for computing embedded boundaries.

data analysis standpoint. By using descriptive statistics to summarize both characteristic features of a data distribution and measures of uncertainty, we can achieve a more cohesive understanding of the information. In this work, we reexamine the box plot and its relatives and develop a new hybrid summary plot that combines moment, cumulant, and density information along with higher order descriptors that rely on distribution fitting. In view of the important role summarizing plots has in decision making, our work focuses on using advanced visualization techniques to incorporate additional descriptive parameters, while simultaneously improving the comprehensibility of summary plots.

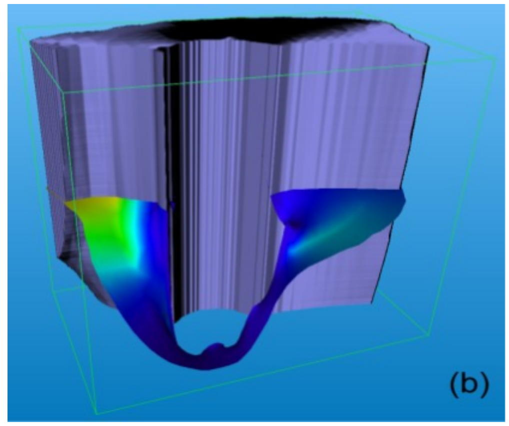
The approach this work takes to understand and visualize uncertainty data focuses on methods for quantitatively displaying the underlying statistics which describe the uncertainty of a dataset. The measures typically used to define uncertainty are mean and variance (standard deviation) and many methods exist in graphical data analysis packages for displaying these quantities. This work extends these methods to include other descriptive statistics as well as distribution fitting techniques to not only provide visualizations of uncertainty, but also give the user a way to better understand the data distribution. Techniques for visualizing the correlation between multiple 1D categorical dataset are also presented (Figure 36), along with a variety of exemplar datasets.

Summer 2008 Internship

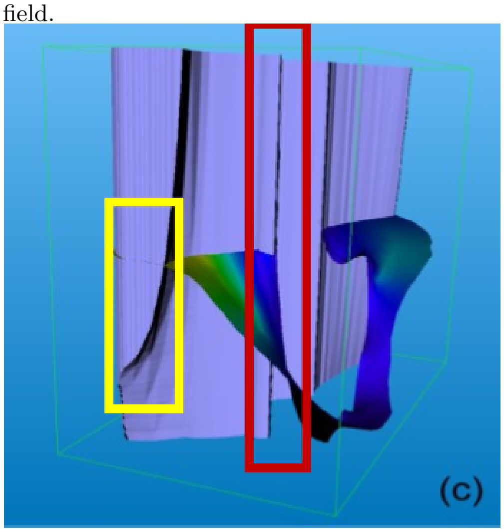
Kristi Potter (Utah) worked at Sandia over the summer, continuing research on uncertainty visualization. The main thrust of this work was to incorporate scientific and information visualization methods into a single system to provide an investigational tool for data exploration. The main focus of the work was on Ensemble data from NOAA, specifically short-term forecast data (SREF), consisting of multiple forecast models run using a variety of input perturbations and forecast hours. The biggest challenges of this work are in the complexity of the data; the numerous variables, sim-



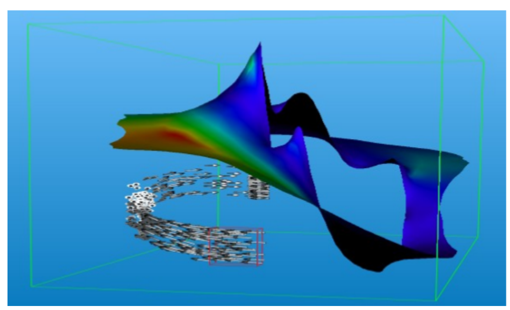
(a) Direct volume rendering of the potentials volume superimposed onto the mean height-field.



(b) Isosurfacing the potentials volume.

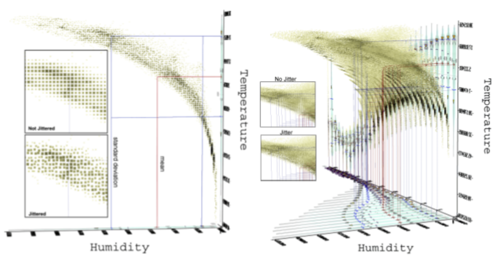


(c) Iso-surfacing of the potentials data. Bends in the iso-surface (yellow box) show areas with high dependence on the input conductivity.

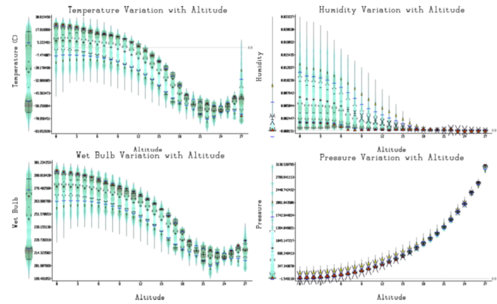


(d) Particle tracing on the potentials volume.

Figure 35: These images show the results of recent research in the field of uncertainty quantification and visualization.



(a) Joint mean (red lines), joint standard deviation (blues lines) and the joint histogram for a single (left) and multiple (right) categorical datasets.



(b) Temperature, humidity, wet bulb, and pressure data taken at a single moment in time across the world. Data courtesy of the Canada Meteorological Centre.

Figure 36: Visualizing statistics using different new techniques.

ulation hours and input perturbations inhibit the direct display of the data and require filtering of some type to intelligibly display. However, the aim of the system is to provide a tool for both hypothesis forming and testing, thus the user must be able to drive the visualization system based to further the exploration of the data. As such, the prototype created over the summer provides a global overview of the data in the form of mean and variation of a specific, user-chosen variable, along with an interactive "drill-down" which shows the actual data values which contribute to the global presentation. The prototype was fully implemented in VTK. Continuing work will extend this approach to other datasets as well as investigate improved visualization techniques.

Future Work

- Research on Direct High Dimension Probabilistic Data Visualization with applications in fusion simulation and biophysical phenomena. Mathematical models used for the reconstruction of experimental results, can produce data distributions with spatial positions in 2 and 3D. These distributions not only express the results of the model, but also fully describe the uncertainty of the results. While much uncertainty visualization research has focused on scalar, or possibly vector values quantifying uncertainty, visualization techniques which expose the entire distribution can lead to a more through understanding. However, the dimensionality of this dataset is too large to be directly displayed or easily comprehended, thus techniques for displaying this data are very important.

- Development of general methods for uncertainty visualization in scalar and vector valued datafields. Most visualization techniques that incorporate uncertainty information treat uncertainty like an unknown or fuzzy quantity. These methods employ the syntax of the word uncertainty to create the interpretation of uncertainty or unknown to indicate areas in a visualization with less confidence, greater error, or high variation. Blurring or fuzzing a visualization, while accurately expressing the lowered confidence one should have in that data, does not lead to a more informative decision making tool, but instead obfuscates the information that lead to the measure of uncertainty. Such a solution to the problem of adding qualitative information to visualization does not elucidate on the quantitative measures leading to the uncertain classification, and thus is missing some important information. The goal of this work is to identify and visualize the measures typically thrown under the umbrella of uncertainty. Uncertainty, as the scientific visualization field titles it refers to quality of a

measured value and can include measures of confidence, error, and deviation. Statistically, uncertainty is harder to identify, since there are many measures that can add to the qualification of data. Using measures that are statistically meaningful to express the uncertainty in a data set exposes insights to the data that may not have previously been obvious. Adding such quantities to visualizations will improve the effectiveness of visualizations by providing a more complete description of the data, and create better tools for decision making and analysis.

- Continue collaboration with Sandia National Lab on the visualization of ensemble datasets. Ensemble datasets consist of multiple mathematical models which predict the result of a simulation. Uncertainty is inherit in these datasets, since the result of each model varies. In addition, sensitivity analyses are often incorporated into the ensembles. Not only is the data itself complex to visualize, the uncertainty associated with this data is of utmost importance. In the upcoming 6 months work will continue on the prototype built in collaboration with SNL to visualize ensemble data, resulting in a submission to the IEEE Visualization conference.

5 Publications, Presentations, Awards, Service, and Outreach

5.1 Publications

5.1.1 Peer-Reviewed Journal Articles

1. Luke J. Gosink, John C. Anderson, E. Wes Bethel, and Kenneth I. Joy. Query-Driven Visualization of Time-Varying Adaptive Mesh Refinement Data. *IEEE Transactions on Visualization and Computer Graphics (Special Issue: Proceedings of IEEE Visualization 2008)*, 14(6), November/December 2008. LBNL-803E.

2. Oliver Rübel, Gunther H. Weber, Min-Yu Huang, E. Wes Bethel, Mark D. Biggin, Charless. C. Fowlkes, C. Luengo Hendriks, Soile. V. E. Keränen, Michael B. Eisen, David W. Knowles, Jitendra Malik, Hans Hagen, and Bernd Hamann. Integrating data clustering and visualization for the analysis of 3d gene expression data. *IEEE Transactions on Computational Biology and Bioinformatics*, 2008. LBNL-382E, to appear.

3. C. Scheidegger, J. Schreiner, B. Duffy, H. Carr, and C. Silva. Revisiting Histograms and Isosurface Statistics. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE Visualization 2008)*, 14(6), 2008.

4. C. Dietrich, J. Comba, L. Nedel, C. Scheidegger, and C. Silva. Edge Groups: A New Approach to Understanding the Mesh Quality of Marching Methods. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE Visualization 2008)*, 14(6), 2008.

5. D. Koop, C. Scheidegger, S. Callahan, J. Freire, and C. Silva. VisComplete: Data-driven Suggestions for Visualization Systems. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE Visualization 2008)*, 14(6), 2008.

6. R. M. Kirby and C. Silva. The Need for Verifiable Visualization. *IEEE Computer Graphics and Applications*, 2008.

7. J. Daniels, C. Silva, J. Shepherd, , and E. Cohen. Quadrilateral Mesh Simplification. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2008)*, 2008.

8. J. Daniels, T. Ochotta, L. Ha, and C. Silva. Spline-Based Feature Curves from Point-Sampled Geometry. *The Visual Computer*, 24(6):449–462, 2008.

9. A. Baptista, B. Howe, J. Freire, D. Maier, and C. Silva. Scientific Exploration in the Era of Ocean Observatories. *Computing in Science and Engineering*, 10(3):53–58, 2008.

10. J. Freire, D. Koop, E. Santos, and C. Silva. Provenance for Computational Tasks: A Survey. *Computing in Science and Engineering*, 10(3):11–21, 2008.

11. E. W. Anderson, J. Ahrens, K. Heitmann, S. Habib, and C. Silva. Provenance in Comparative Analysis: A Study in Cosmology. *Computing in Science and Engineering*, 10(3):30–37, 2008.

12. L. Bavoil, S. Callahan, and C. Silva. Robust Soft Shadow Mapping with Depth Peeling. *Journal of Graphics Tools*, 13(1):19–30, 2008.

13. C. Scheidegger, D. Koop, E. Santos, H. Vo, S. Callahan, J. Freire, and C. Silva. Tackling the Provenance Challenge One Layer at a Time. *Concurrency And Computation: Practice And Experience*, 20(5):473–483, 2008.

14. S. P. Callahan, J. H. Callahan, C. E. Scheidegger, and C. Silva. Direct Volume Rendering: A 3D Plotting Technique for Scientific Data. *Computing in Science and Engineering*, 10(1):88–92, 2008.

15. C. Silva and J. Tohline. Guest Editorial: Special Issue on Computational Provenance. *Computing in Science and Engineering*, 10(3):9–10, 2008.

16. Gunther H. Weber, Oliver Rübel, Min-Yu Huang, Anagla H. DePace, Charless C. Fowlkes, Soile V.E. Keränen, Cristian L. Luengo Hendriks, Hans Hagen, David W. Knowles, Jitendra Malik, Mark D. Biggin, and Bernd Hamann. Visual Exploration of Three-dimensional Gene Expression using Physical Views and Linked Abstract Views. *IEEE Transactions on Computational Biology and Bioinformatics*, 2008. To appear. LBNL-6377.

17. Cecilia Aragon, Stephen Bailey, Sarah Poon, Karl Runge, and Rollin Thomas. Sunfall: A Collaborative Visual Analytics System for Astrophysics. *Journal of Physics Conference Series – SciDAC 2008*, 125, June 2008. LBNL-657E.

18. C. C. Fowlkes, C. L. Luengo Hendriks, S. V. E. Keränen, G. H. Weber, O. Rübel, M.-Y. Huang, S. Chatoor, A. H. DePace, L. Simirenko, C. Henriquez, A. Beaton, R. Weiszmann, S. Celniker, B. Hamann, D. W. Knowles, M. D. Biggin, M. B. Eisen, and J. Malik. A Quantitative Spatio-temporal Atlas of Gene Expression in the Drosophila Blastoderm. *Cell*, 133:364–374, April 2008.

19. Brian Paul, Sean Ahern, E. Wes Bethel, Eric Brugger, Rich Cook, Jamison Daniel, Ken Lewis, Jens Owen, and Dale Southard. Chromium Renderserver: Scalable and Open Remote Rendering Infrastructure. *IEEE Transactions on Visualization and Computer Graphics*, 14(3), May/June 2008. LBNL-63693.

20. John C. Anderson, Christoph Garth, Mark A. Duchaineau, and Kenneth I. Joy. Discrete Multi-Material Interface Reconstruction for Volume Fraction Data. *Computer Graphics Forum (Proc. of Eurographics/IEEE-VGTC Symposium on Visualization 2008)*, 27(3):1015–1022, 2008.

21. Christoph Garth, Xavier Trichoche, and Kenneth I. Joy. Lagrangian Visualization of Flow-Embedded Surface Structures. *Computer Graphics Forum (Proc. of Eurographics/IEEE-VGTC Symposium on Visualization 2008)*, 27(3):767–774, 2008.

22. R. Westermann K. Buerger, J. Krueger. Direct Volume Editing. *IEEE Transactions on Visualization and Computer Graphics (Special Issue: Proceedings of IEEE Visualization 2008)*, 14(6), November/December 2008.

23. M. Chen, M. Kirby, C. Hansen, and T. Moller. A Staggered Grid for Multi-resolution Volume Rendering. *Computer Graphics Forum Journal*, Accepted, to appear.

24. Aaron Knoll, Younis Hijazi, Charles Hansen, Ingo Wald, and Hans Hagen. Interactive Ray Tracing of Arbitrary Implicit Functions. *Computer Graphics Forum Journal*, Accepted, to appear.

25. Aaron Knoll, Ingo Wald, and Charles Hansen. Coherent Multiresolution Isosurface Ray Tracing. *The Visual Computer*, 2008. To Appear.

26. Guo-Shi Li, Charles Hansen, and Xavier Tricoche. Physically-based Dye Advection for Flow Visualization. *Computer Graphics Forum Journal*, 27(3):727–734, 2008.

27. Guo-Shi Li, Xavier Tricoche, Daniel Weiskopf, and Charles Hansen. Flow Charts: Visualization of Vector Fields on Arbitrary Surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 14(5):1067–1080, 2008.

28. Jens Bellmann, Frank Michel, Eduard Deines, Martin Hering-Bertram, Jan Mohring, and Hans Hagen. Sound Tracing: Rendering Listener Specific Acoustic Room Properties. *Computer Graphics Forum*, 27(3):943–950, 2008.

5.1.2 Conference Proceedings

1. Oliver Rübel, Prabhat, Kesheng Wu, Hank Childs, Jeremy Meredith, Cameron G. R. Geddes, Estelle Cormier-Michel, Sean Ahern, Gunther H. Weber, Peter Messmer, Hans Hagen, Bernd Hamann, and E. Wes Bethel. High Performance Multivariate Visual Data Exploration for Extemely Large Data. In *SuperComputing 2008 (SC08)*, Austin, Texas, USA, November 2008. LBNL-716E (to appear).

2. Daniela M. Ushizima, Oliver Rübel, Prabhat, Gunther H. Weber, E. Wes Bethel, Cecilia R. Aragon, Cameron G.R. Geddes, Estelle Cormier-Michel, and Bernd Hamann. Automated Analysis for Detecting Beams in Simulations. In *Proceedings of the Seventh International Conference on Machine Learning and Applications*, December 2008. LBNL-960E.

3. H. Wang, C. E. Scheidegger, and C. Silva. Optimal Bandwidth Selection for MLS Surfaces. In *IEEE International Conference on Shape Modeling and Applications (SMI) 2008*, 2008.

4. E. Santos, L. Lins, J. P. Ahrens, J. Freire, and C. Silva. A First Study on Clustering Collections of Workflow Graphs. In *Second International Provenance and Annotation Workshop (IPAW 2008)*, 2008.

5. S. P. Callahan, J. Freire, C. E. Scheidegger, C. Silva, and Huy T. Vo. Towards Provenance-Enabling ParaView. In *Second International Provenance and Annotation Workshop (IPAW 2008)*, 2008.

6. T. Ellkvist, D. Koop, E. W. Anderson, J. Freire, and C. Silva. Using Provenance to Support Real-Time Collaborative Design of Workflows. In *Second International Provenance and Annotation Workshop (IPAW 2008)*, 2008.

7. L. Lins, D. Koop, E. W. Anderson, S. P. Callahan, E. Santos, C. E. Scheidegger, J. Freire, and C. T. Silva. Examining Statistics of Workflow Evolution Provenance: A First Study. In *Statistical and Scientific Database Management (SSDBM) 2008*, 2008.

8. C. E. Scheidegger, H. T. Vo, D. Koop, J. Freire, and C. Silva. Querying and Re-Using Workflows with VisTrails. In *ACM SIGMOD 2008*, 2008.

9. J. Freire and C. Silva. Towards Enabling Social Analysis of Scientific Data. In *CHI Social Data Analysis Workshop 2008*, 2008.

10. Cecilia Aragon, Sarah Poon, Gregory Aldering, Rollin Thomas, and Robert Quimby. Using Visual Analytics to Maintain Situational Awareness in Astrophysics. In *Proceedings of 2008 IEEE Symposium on Visual Analytics Science and Technology.* IEEE Computer Society Press, October 2008. LBNL-658E.

11. Gunther H. Weber, Vincent E. Beckner, Hank Childs, Terry J. Ligocki, Mark Miller, Brian van Straalen, and E. Wes Bethel. Visualization of Scalar Adaptive Mesh Refinement Data. 385:309–320, 2008. LBNL-220E.

12. David Pugmire, Hank Childs, and Sean Ahern. Parallel Analysis and Visualization on Cray Compute Node Linux. In *Cray Users Group Meeting*, Helsinki Finland, May 2008.

13. K. Potter, J. Krueger, and C.R. Johnson. Towards the Visualization of Multi-Dimensional Stochastic Distribution Data. In *Proceedings of The International Conference on Computer Graphics and Visualization (IADIS) 2008*, 2008.

5.1.3 Invited Articles

1. C. Silva and J. Freire. Software Infrastructure for Exploratory Visualization and Data Analysis: Past, Present and Future. In *Journal of Physics: Conference Series (SciDAC 2008 Conference)*, 2008.

2. E. Wes Bethel, Chris Johnson, Charles Hansen, Claudio Silva, Steven Parker, Allen Sanderson, Lee Myers, Martin Cole, Xavier Tricoche, Sean Ahern, George Ostrouchov, Dave Pugmire, Jamison Daniel, Jeremy Meredith, Valerio Pascucci, Hank Childs, Peer-Timo Bremer, Ajith Mascarenhas, Ken Joy, Bernd Hamann, Christoph Garth, Cecilia Aragon, Gunther Weber, and Prabhat. Seeing the Unseeable. *SciDAC Review*, (8):24–33, Summer 2008. LBNL-472E.

3. Oliver Rübel, Gunther H. Weber, Min-Yu Huang, E. Wes Bethel, Soile V. E. Keränen, Charless C. Fowlkes, Cris L. Luengo Hendriks, Angela H. DePace, Lisa. Simirenko, Michael B. Eisen, Mark D. Biggin, Hans Hagen, Jitendra Malik, David W. Knowles, and Bernd Hamann. *PointCloudXplore 2: Visual Exploration of 3D Gene Expression.* GI Lecture Notes in Informatics. Gesellschaft fuer Informatik (GI), 2008. LBNL-249E.

5.1.4 Book Chapters

1. E. Wes Bethel, Hank Childs, Ajith Mascarenhas, Valerio Pascucci, and Prabhat. Scientific Data Managment Challenges in High Performance Visual Data Analysis. In Arie Shoshani and Doron Rotem, editors, *Scientific Data Management: Challenges, Existing Technology, and Deployment*. Chapman & Hall/CRC Press, 2008. (to appear).

5.1.5 Posters

1. Oliver Rübel, Prabhat, Kesheng Wu, Hank Childs, Jeremy Meredith, Cameron G. R. Geddes, Estelle Cormier-Michel, Sean Ahern, Gunther H. Weber, Peter Messmer, Hans Hagen, Bernd Hamann, and E. Wes Bethel. Application of High-performance Visual Analysis Methods to Laser Wakefield Particle Acceleration Data. In *IEEE Visualization 2008*, Columbus, Ohio, USA, October 2008. (LBNL report number pending) (to appear).

2. E. Wes Bethel, C.R. Johnson, C. Aragon, Prabhat, O. Rübel, G. Weber, V. Pascucci, H. Childs, P.-T. Bremer, A. Mascarenhas, B. Whitlock, S. Ahern, J. Meredith, G. Ostrouchov, K. Joy, B. Hamann, C. Garth, M. Cole, C. Hansen, S. Parker, A. Sanderson, C.T. Silva, and X. Tricoche. DOE SciDAC Visualization and Analytics Center for Enabling Technologies. In *2008 DOE ASCR CS PI Meeting*, Denver, CO, USA, April 2008.

5.1.6 Technical Reports

1. Luke J. Gosink, Kesheng Wu, E. Wes Bethel, John D. Owens, and Kenneth I. Joy. Bin-Hash Indexing: A Parallel Method for Fast Query Processing. Technical Report LBNL-729E, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, 94720, 2008.

5.1.7 Recently Submitted Publications

1. K. Potter, J. Kniss, R. Riesenfeld, and C. Johnson. Visualizing Summary Statistics and Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 2008. Fast-tracked from IEEE VIS submission.

5.2 Presentations

5.2.1 Invited Presentations

1. E. Wes Bethel. High Performance, Query-Driven Scientific Visualization: Finding Smaller Needles in Larger Haystacks. In *University of Tulsa Research Seminar*, Tulsa, OK, USA, September 2008.

2. E. Wes Bethel. Parallelism in Graphics and Visualization. In *University of Tulsa, CS6813 Guest Lecture*, Tulsa, OK, USA, September 2008.

3. E. Wes Bethel. Accelerating Visual Knowledge Discovery with Query-Driven Visualization. In *SIAM Conference on Imaging Science (IS08), Visualization and Analytics for Science Discovery*, San Diego, CA, USA, July 2008.

4. E. Wes Bethel. Scientific Visualization: The Modern Oscilloscope for Seeing the Unseeable. In *LBNL Summer Lecture Series*, Berkeley, CA, USA, June 2008. Also on YouTube at `http://www.youtube.com/watch?v=R4LLuEOHTtE`.

5. E. Wes Bethel. Modern Scientific Visualization is More Than Just Pretty Pictures. In *Numerical Modeling of Space Plasma Flows: Astronum-2008 (Astronomical Society of the Pacific Conference Series)*, St. John, USVI, June 2008.

6. Gunther H. Weber. Visualization and Analysis of Adaptive Mesh Refinement Data with VisIt. In *Numerical Modeling of Space Plasma Flows: Astronum-2008 (Astronomical Society of the Pacific Conference Series)*, St. John, USVI, June 2008.

7. E. Wes Bethel. Visual Data Analysis and Data Exploration at the Extreme Scale. In *DOE Office of Science, Advanced Scientific Computing Research Principal Investigator Meeting*, Denver, CO, USA, April 2008.

8. Hank Childs. Why Petascale Visualization Will Change the Rules. In *International Conference on Computational Science 2008 (ICCS 2008)*, Krakow, Poland, June 2008. Keynote Presentation.

9. Hank Childs. Why Petascale Visualization Will Change the Rules. In *Center for Scalable Application Development Software (CScADS) Summer Workshop on Scientific Data Analysis and Visualization for Petascale Computing*, Snowbird, Utah, USA, July 2008.

10. Hank Childs. Petascale Visualization and VisIt. In *Colloquium Series for the Center for Computation and Technology (CCT) at Louisiana State University*, Baton Rouge, LA, USA, September 2008.

11. Gunther H. Weber. Visualization Tools for Adaptive Mesh Refinement Data. In *DOE Computer Graphics Forum*, Duck, NC, USA, April 2008.

12. David Pugmire, Hank Childs, and Sean Ahern. Parallel Analysis and Visualization on Cray Compute Node Linux. In *DOE Computer Graphics Forum*, Duck, NC, USA, April 2008.

13. David Pugmire, Hank Childs, and Sean Ahern. Parallel Analysis and Visualization on Cray Compute Node Linux. In *Cray Users Group Meeting*, Helsinki Finland, May 2008.

14. Charles Hansen. Multidimensional Transfer Functions and other GPU Methods. In *Exxon-Mobile*, Houston, TX, USA, August 2008.

15. Claudio Silva. Unstructured Grids and High-Quality Surface Reconstruction from Different Data Types. In *Exxon-Mobile*, Houston, TX, USA, August 2008.

16. Claudio Silva. VisTrails: Provenance and Data Exploration. In *National Biomedical Computation Resource (NBCR) Summer Institute*, San Diego, CA USA, August 2008.

17. Claudio Silva. Software Infrastructure for Exploratory Visualization and Data Analysis: Past, Present and Future. In *SciDAC 2008*, Seattle, WA USA, July 2008.

18. Charles Hansen. Interactive Texture-based Flow Visualization. In *LANL*, Los Alamos, NM, USA, August 2008.

19. Charles Hansen. CSAFE. In *University of Kaiserslautern Colloquium*, Kaiserslautern, GERMANY, May 2008.

20. George Ostrouchov. Data-Parallel Analysis and Graphics with R. In *DOE Computer Graphics Forum*, Duck, NC, USA, April 2008.

21. George Ostrouchov. Stalking the Interactive Terabyte with R: Data-Parallel Statistical Computing. In *University of Tennessee SOMS Seminar Series*, Knoxville, TN, USA, April 2008.

5.3 Tutorials

5.3.1 VisIt Tutorials

1. SciDAC 2008 Program Meeting, June 2008, Seattle WA. Presenters: Hank Childs and Sean Ahern. Approximately 20 attendees.
2. CScADS workshop, July 2008, Snowbird Utah. Presenters: Hank Childs and Jeremy Meredith. Approximately 20 attendees.
3. Princeton Plasma Physics Laboratory, September 2008. Presenter: Sean Ahern. Approximately 20 attendees.

5.3.2 VisTrails Tutorials

1. SciDAC 2008 Program Meeting, June 2008, Seattle WA. Presenters: Claudio Silva and Carlos Scheidegger. Approximately 20 attendees.
2. CScADS workshop, July 2008, Snowbird Utah. Presenter: Claudio Silva. Approximately 20 attendees.

Figure 37: Sean Ahern delivers a VisIt tutorial to fusion researchers at the Princeton Plasma Physics Laboratory in September 2008.

5.4 Workshops

1. SciDAC Framework Application for Core-Edge Transport Simulations (FACETS) Fall 2008 Program meeting. September 4-5, 2008. Dave Pugmire, Allen Sanderson.

2. Center for Scalable Application Development Software (CScADS) workshop on Scientific Data Analysis and Visualization for Petascale Computing. Hank Childs, Jeremy Meredith.

3. Office of Nuclear Energy's NEAMS program (Nuclear Energy Advanced Modeling and Simulation) Workshop on Enabling Technologies. Hank Childs, Wes Bethel. Childs chaired breakout session on results.

4. DoE Mathematics for the Analysis of Petascale Data Workshop (MAPD). Kenneth I. Joy and Valerio Pascucci.

5. Third Annual Workshop of the International Research Training Group (IRTG). Kaiserslautern, September 29 - November 1, 2008. Participants: Kenneth I. Joy, Christoph Garth, and Eduard Deines.

5.5 Awards

1. Best Paper Award. Optimal Bandwidth Selection for MLS Surfaces. H. Wang, C. E. Scheidegger, and C. Silva. IEEE International Conference on Shape Modeling and Applications (SMI) 2008..

2. SciDAC 2008 OASCARS. VACET researchers win three "People's Choice Awards" at Visualization Night at the SciDAC 2008 Program meeting in Seattle WA in June 2008, as well as one "Honorable Mention. These are listed and described in more detail here: `http://www.vacet.org/news/news2008.html#scidac`. See Figure 38.

3. Stakeholders win OASCARS at SciDAC 2008. VACET science stakeholders at Tech-X Corporation win an OASCAR at the SciDAC 2008 meeting using VACET software to show off their VORPAL code, which is used in high energy physics and fusion applications.

4. NVIDIA Recognized University of Utah as a CUDA Center of Excellence.

5.6 Service

5.6.1 Technical Reviewer

1. IEEE Transactions on Visualization and Computer Graphics. Associate Editor: V. Pascucci. Technical Paper Reviewer: Wes Bethel, Gunther H. Weber. Kenneth I. Joy, Peer-Timo Bremer, Jens Krüger, Kristi Potter, Charles Hansen, Valerio Pascucci, Christoph Garth.

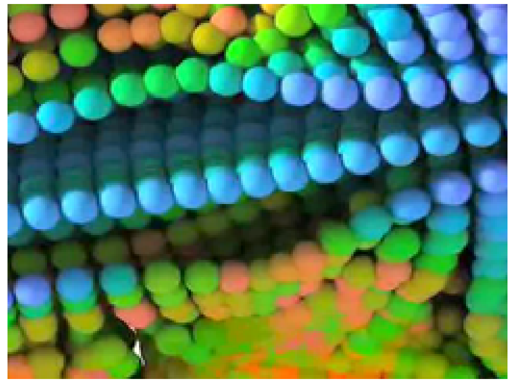
2. IEEE Visualization 2008. Technical Paper Reviewer: Wes Bethel, Hank Childs, Gunther H. Weber, Kenneth I. Joy, Peer-Timo Bremer, Jens Krüger, Kristi Potter, Valerio Pascucci, Christoph Garth.

3. IEEE VAST 2008. Technical paper reviewer: Gunther H. Weber.

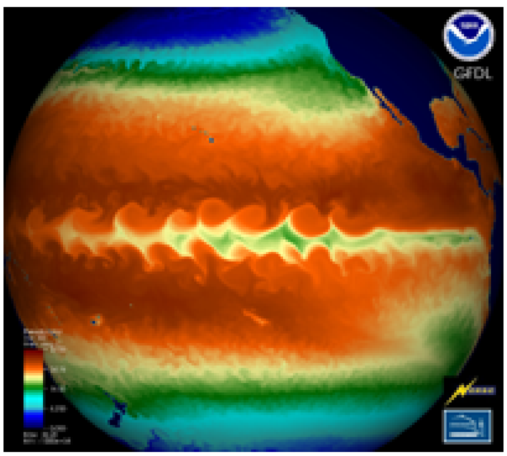
4. DOE SBIR/STTR Technical Proposal Reviewer: E. Wes Bethel.

5. Journal of Systems and Software: Gunther H. Weber.

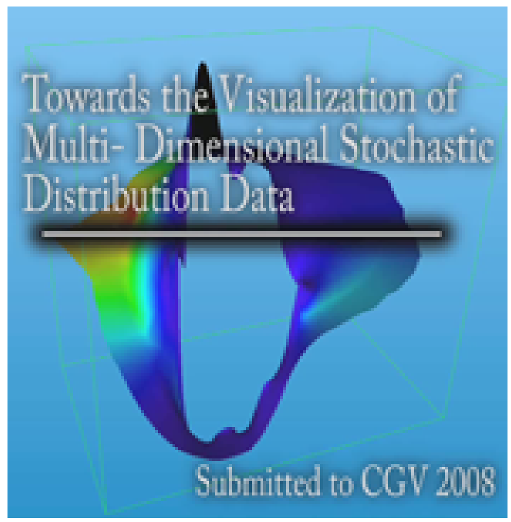
(a) Accurate and Efficient Integral Surfaces for Flow Visualization. C. Garth, H. Krishnana, X. Tricoche, and K. Joy.



(b) The Golden Age of Scientific Visualization. C. Johnson et al.



(c) Visualization of Next-generation GFDL/NOAA Climate Simulations. Prabhat, J. Jacobsen, G. Weber, C. Kerr, V. Balaji, W. Bethel..



(d) Towards the Visualization of Multi-Dimensional Stochastic Distribution Data. K. Potter, J. Krüger, C. Johnson.

Figure 38: VACET researchers with three “People’s Choice Awards” and one “honorable mention” at Visualization Night at the SciDAC 2008 Program meeting in Seattle, WA.

6. ACM Transactions on Graphics. Technical Paper Reviewers: Peer-Timo Bremer, Jens Krüger, Valerio Pascucci.

5.6.2 Program Committee

1. ACM Multimedia 2008 Technical Demonstrations: Claudio Silva.
2. Knowledge-Assisted Visualization (KAV) 2008: Claudio Silva.
3. International Symposium on Volume Graphics 2008 (VG08): Claudio Silva.
4. XIX Brazilian Symp on Computer Graphics and Image Processing (SIBGRAPI) 2008: Claudio Silva.

FOR IMMEDIATE RELEASE:

NVIDIA RECOGNIZES UNIVERSITY OF UTAH AS A CUDA CENTER OF EXCELLENCE

University of Utah Latest in a Growing List of Exceptional Schools Demonstrating Pioneering Work in Parallel Computing

SANTA CLARA, CA & SALT LAKE CITY, UT —JULY 31, 2008—NVIDIA Corporation, the worldwide leader in visual computing technologies, and the University of Utah today announced that the university has been recognized as a CUDA Center of Excellence, a milestone that marks the beginning of a significant partnership between the two organizations.

5. Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT) 2008: Claudio Silva.
6. 2nd International Provenance and Annotation Workshop (IPAW 2008): Claudio Silva.
7. ACM SIGGRAPH 2008 Papers Program: Claudio Silva.
8. ACM Solid and Physical Modeling Symposium (SPM) 2008: Claudio Silva.
9. IEEE International Conference on Shape Modeling and Applications (SMI) 2008: Claudio Silva.
10. EuroVis 2008: Claudio Silva.
11. International Conference on Computer Animation and Social Agents (CASA) 2008: Claudio Silva.
12. Symposium on Geometry Processing 2008: Claudio Silva.
13. Symposium on Computational Geometry 2009. General Conference co-Chair: V. Pascucci.
14. IEEE/Eurographics International Symposium on Volume Graphics 2008. General Conference co-Chair: V. Pascucci.
15. SIAM invited minisymposium on "Visualization and Analytics for Science Discovery¡i¿ (at the SIAM Annual Meeting, 2008). Chair: V. Pascucci.
16. SIAM Conference on Data Mining 2008: George Ostrouchov.
17. DOECGF Steering Committee: Wes Bethel and Sean Ahern.
18. IEEE Workshop Knowledge-assisted Visualization: Gunther H. Weber.
19. TopoInVis 2009. Organization Committee: Valerio Pascucci; Program Committee: Peer-Timo Bremer, Gunther H. Weber, Charles Hansen, and Christoph Garth.
20. IEEE Visualization 2008. Papers Co-Chair: Charles Hansen. General Conference co-Chair: Kenneth I. Joy. Exhibits Chair: Kenneth I. Joy. Papers Committee: Kenneth I. Joy.

21. Information and Knowledge Sciences Review Committee, Los Alamos National Laboratory: Kenneth I. Joy.
22. EUROVIS 2008: European Visualization Conference 2008: Kenneth I. Joy and Charles Hansen.
23. EUROGRAPHICS 2009. General Areas co-Chair: Jens Krüger. Program Committee: V. Pascucci.
24. Eurographics 2008 Symposium on Parallel Graphics and Visualization (EGPGV '08): Charles Hansen and Valerio Pascucci.
25. Joint Statistical Meetings 2010. Elected Program Chair for section on Physical and Engineering Sciences: George Ostrouchov.

5.7 VACET Website

VACET expends a great deal of effort to keep its website, `http://www.vacet.org`, up to date. This period, we have:

- Reorganized our Images/Movies gallery pages, and added a substantial amount of content.
- Added news items as they occur.
- Updated our publications page, as well as maintain a BibTex file of VACET publications.
- Maintained a "Software" page where visitors can download the latest version of our software applications.

6 Resources

6.1 NERSC/LBNL

VACET applied for and received an ERCAP allocation at NERSC for AY08. We were awarded 50,000 CPU hours, approximately 50TB of archival storage, and approximately 1TB on the NERSC global filesystem (GFS). Recently, we requested and received a temporary GFS quota increase to 5TB in order to support "hero-sized" visual data analysis projects in support of our science stakeholders in accelerator and combustion. That effort, conducted jointly with our stakeholders, produced novel research results as well as multiple paper submissions to IEEE Transactions on Visualization and Computer Graphics and Supercomputing 2008.

During the course of this AY, we ended up consuming way more cycles than originally forecast. The bulk of this overrun resulted from the cost of doing parallel I/O performance for the Climate/Randall project (Section 2.6). The ERCAP allocation manager, Francesca Verdier, made several additions of cycles to our allocation over the course of the project – we are grateful for her help.

We will be completing a new ERCAP allocation for resources at NERSC. The application is due 10/1/2008.

6.2 LCF/ORNL

At LCF/ORNL, VACET has access to Jaguar through a Director's Discretionary Allocation. Additionally, we have ongoing access to ORNL's visualization cluster.